



Departamento de Computação
Curso de Bacharelado em Ciências da Computação

Thiago Luiz Barbosa da Silva

Um estudo de caso para previsão de partidas de futebol utilizando o ChatGPT

Recife

2024

Thiago Luiz Barbosa da Silva

Um estudo de caso para previsão de partidas de futebol utilizando o ChatGPT

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Ciências da Computação da Universidade Federal Rural de Pernambuco como requisito para obtenção do título de Bacharel em Ciências da Computação.

Orientador: Prof. Dr. Leandro Marques do Nascimento.

Recife

2024

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Auxiliadora Cunha – CRB-4 1134

S586e Silva, Thiago Luiz Barbosa da.
Um estudo de caso para previsão de partidas de futebol utilizando o ChatGPT / Thiago Luiz Barbosa da Silva. - Recife, 2024.
51 f.; il.

Orientador(a): Leandro Marques do Nascimento.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural de Pernambuco, Bacharelado em Ciência da Computação, Recife, BR-PE, 2024.

Inclui referências.

1. Inteligência artificial. 2. Chatgpt. 3. Large Language Model (LLM) . 4. Futebol 5. Resultados. I. Nascimento, Leandro Marques do, orient. II. Título

CDD 004

Thiago Luiz Barbosa da Silva
Um estudo de caso para previsão de partidas de futebol utilizando o
ChatGPT

Trabalho de conclusão de curso apresentado ao
Curso de Bacharelado em Ciências da Computa-
ção da Universidade Federal Rural de Pernambuco,
para obtenção do título de Bacharel em Ciências da
Computação.

Aprovado em: 01/10/2024

BANCA EXAMINADORA

Leandro Marques do Nascimento - Doutor
Universidade Federal Rural de Pernambuco

George Augusto Valença dos Santos - Doutor
Universidade Federal Rural de Pernambuco

Dedico este trabalho aos meus familiares e amigos, cuja paciência, incentivo e compreensão foram essenciais em cada etapa deste processo. Àqueles que, de forma direta ou indireta, contribuíram para o meu crescimento pessoal e acadêmico, sou eternamente grato. Agradeço, especialmente, aos meus professores e colegas, cujas trocas de conhecimento e experiência foram fundamentais para a realização deste estudo.

Agradecimentos

Gostaria de expressar minha profunda gratidão a todos que, de diferentes maneiras, contribuíram para a realização deste trabalho. Agradeço à minha família, por todo apoio, carinho e compreensão durante as fases mais desafiadoras desta jornada. Aos amigos, por sempre estarem presentes com palavras de encorajamento e momentos de descontração que ajudaram a manter o equilíbrio.

Por fim, deixo meu agradecimento especial ao meu orientador, Leandro Marques, cuja orientação, paciência e sabedoria foram essenciais para a conclusão deste trabalho. Sua dedicação e entusiasmo pela pesquisa me inspiraram a buscar sempre o melhor. Muito obrigado!

Resumo

O presente trabalho tem como objetivo desenvolver e testar uma ferramenta de previsão de resultados de partidas de futebol, utilizando o modelo de linguagem ChatGPT. A pesquisa explora o potencial dessa tecnologia para processar dados de partidas e gerar previsões, comparando seu desempenho com as probabilidades oferecidas por casas de apostas. O método utilizado inclui a coleta de dados por meio de web scraping em fontes como Placar de Futebol e FBref, o que possibilitou a criação de uma base de dados rica em informações sobre equipes, campeonatos e estatísticas detalhadas. A partir dessa base, a ferramenta foi criada dentro do projeto Arena Sport Club, que possui funcionalidades sobre visualização de resultados e informações sobre futebol. Diferentes estratégias de geração de *prompts* foram implementadas na ferramenta para verificar a melhor maneira de instruir o modelo a fornecer previsões precisas. Os resultados mostraram que o modelo tem potencial para realizar previsões de resultados de futebol de maneira eficaz, aproximando-se das taxas de acerto das casas de apostas. Entretanto, o trabalho identificou desafios, como o alto custo financeiro e a necessidade de ajustes contínuos para lidar com a complexidade das partidas e as variáveis envolvidas. A conclusão sugere que, embora o ChatGPT ofereça uma ferramenta promissora para previsões esportivas, é necessário otimizar seu uso em contextos reais. Futuras pesquisas podem aprimorar a aplicação dessa tecnologia, reduzindo custos e melhorando a precisão em longo prazo.

Palavras-chave: inteligência artificial; chatgpt; large language model (llm); futebol; resultados.

Abstract

The present study aims to develop and test a tool for predicting football match outcomes using the ChatGPT language model. The research explores the potential of this technology to process match data and generate predictions, comparing its performance with the probabilities offered by betting houses. The method includes data collection through web scraping from sources such as Placar de Futebol and FBref, which allowed the creation of a rich database with detailed information about teams, championships, and statistics. From this database, the tool was developed within the Arena Sport Club project, which includes features for visualizing results and football-related information. Different prompt-generation strategies were implemented in the tool to determine the best way to instruct the model to provide accurate predictions. The results showed that the model has the potential to make effective football match predictions, approaching the accuracy rates of betting houses. However, the study identified challenges such as high financial costs and the need for continuous adjustments to address the complexity of the matches and the variables involved. The conclusion suggests that while ChatGPT offers a promising tool for sports predictions, its use in real-world contexts needs to be optimized. Future research can enhance the application of this technology, reducing costs and improving accuracy over time.

Keywords: artificial intelligence; chatgpt; large language model (llm); football; results.

Sumário

1	INTRODUÇÃO	10
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Evolução dos LLMs e surgimento do ChatGPT	14
2.2	Prompt Engineering	15
2.3	Restrições do ChatGPT	16
2.4	Aplicação de IA no futebol	16
2.5	Trabalhos relacionados	17
3	METODOLOGIA	20
3.1	Seleção do LLM	21
3.2	Obtenção dos dados	23
3.3	Funcionalidade de previsão do resultado de partidas	26
3.3.1	Categorização dos dados para previsão	27
3.3.2	Elaboração dos prompts	30
3.3.2.1	Técnicas utilizadas nos prompts desenvolvidos	31
3.3.2.2	Utilização dos prompts na API do ChatGPT	32
3.4	Exemplo da funcionalidade de previsão do resultado de partidas	32
4	RESULTADOS	35
4.1	Estruturação dos testes	35
4.2	Combinações de testes	36
4.2.1	Combinação simples entre duas categorias	38
4.2.2	Combinações da categoria de temperatura	39
4.2.3	Combinações da categoria de número de exemplos para o modelo	40
4.2.4	Combinações da categoria de número de partidas no histórico	40
4.2.5	Combinações da categoria de estatísticas do campeonato	41
4.2.6	Combinações da categoria do modelo	42
4.3	Comparações com os resultados gerais	43
4.4	Custos por requisição à API do ChatGPT	44
4.5	Dispersão média dos resultados	44
5	CONCLUSÃO E DISCUSSÃO DOS RESULTADOS	46
5.1	Discussão dos resultados obtidos	46
5.2	Pontos fortes, fracos e limitações dos resultados obtidos	49

REFERÊNCIAS	51
--------------------------	-----------

1 Introdução

Após o lançamento do ChatGPT (BANG et al., 2023) no final de 2022, observou-se uma adoção crescente de ferramentas baseadas em inteligência artificial (IA), seja por mera curiosidade ou como auxílio em tarefas cotidianas e profissionais. Em abril de 2023, apenas cinco meses após o seu lançamento, o ChatGPT já acumulava aproximadamente 1,8 bilhões de visitas (DUARTE, 2023). Esse crescimento exponencial no número de acessos despertou significativa atenção e interesse. Em pouco tempo, tornou-se comum o uso dessas ferramentas para pesquisas simples, que anteriormente eram exclusivamente dirigidas a motores de busca tradicionais. A principal diferença reside na capacidade de fornecer respostas mais precisas e diretas, além de apresentarem uma interface mais intuitiva e atraente para todas as faixas etárias. Qualquer pessoa familiarizada com aplicativos de mensagens pode facilmente se adaptar à interface dessas ferramentas.

Assim como outras ferramentas, o ChatGPT utiliza um Modelo de Linguagem de Larga Escala (*Large Language Model* - LLM) (BOMMASANI et al., 2022). Cada LLM possui características e objetivos próprios, diferenciando-se principalmente na forma como são gerados e treinados (BROWN et al., 2020). Esses modelos empregam diversas técnicas de IA e uma quantidade massiva de dados, que são coletados e armazenados para o treinamento do modelo (DAI et al., 2019). Esses dados podem incluir textos, áudios, imagens, entre outros tipos de informações. No caso do ChatGPT, ele utiliza o *Generative Pre-training Transformer* (GPT), com 570 GB de dados e aproximadamente 175 bilhões de parâmetros em seu treinamento (BROWN et al., 2020). Esse modelo aplica diversos algoritmos e técnicas de IA, destacando-se o processamento de linguagem natural (*Natural Language Processing* - NLP) (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011) e redes neurais (WU; FENG, 2018), para compreender perguntas recebidas e fornecer respostas em linguagem natural.

Para alcançar um nível satisfatório de precisão, é necessária uma grande quantidade e variedade de dados. Entretanto, à medida que se aumenta o treinamento, cresce também a necessidade de poder computacional. As grandes empresas de IA buscam aprimorar esses treinamentos, tornando-os menos custosos e mais eficientes em larga escala. O ChatGPT é uma das ferramentas notáveis que já apresenta resultados impressionantes provenientes de treinamento com uma quantidade significativa de dados (BROWN et al., 2020). O sucesso do ChatGPT demonstra como a tecnologia pode ser aplicada em diversos campos para resolver problemas complexos, inclusive além do âmbito digital puro.

Um exemplo claro dessa aplicação transversal é visto na indústria esportiva, particularmente no futebol, um esporte globalmente renomado que engloba diversas características para seu funcionamento eficaz. Entre essas características, destaca-se a área de análises e estatísticas, cada vez mais essencial no contexto esportivo. Segundo a American Statistical Association (ASA), desde a criação de uma seção dedicada às estatísticas desportivas, em 1992, a utilização de estatísticas técnicas nos esportes foi promovida a outro nível (IEVOLI; PALAZZO; RAGOZINI, 2021). Aprofundar-se na análise do adversário por meio de estudos e dados estatísticos permite que as equipes se preparem de maneira mais abrangente para as partidas, antecipando situações específicas e, conseqüentemente, ampliando as possibilidades de vitória. Neste contexto, o uso de algoritmos como a regressão logística, conforme discutido no artigo (PRASETIO; HARLILI, 2016), exemplifica como modelos matemáticos podem prever os resultados das partidas com base em variáveis históricas e situacionais. A regressão logística, em particular, é aplicada para classificar resultados em categorias discretas (vitória, derrota ou empate), utilizando uma função logística para estimar a probabilidade de cada desfecho.

Em contrapartida, os LLMs, como o ChatGPT, empregam redes neurais profundas para gerar previsões baseadas não apenas em dados numéricos, mas em uma vasta gama de informações contextuais e linguísticas. Essa abordagem permite uma análise mais nuanciada das dinâmicas de jogo, oferecendo *insights* que vão além das estatísticas convencionais. Com o tempo, aplicativos de resultados de partidas de futebol facilitaram a visualização dessas estatísticas para os torcedores, difundindo ainda mais o uso desses dados e o interesse por estatísticas mais precisas e diversificadas.

Dito isso, o projeto Arena Sport Club¹ é um projeto acadêmico que foi inicialmente desenvolvido na disciplina de Desenvolvimento de Aplicações Web no final de 2022. O projeto possui funcionalidades principais relacionadas aos resultados de partidas de futebol, incluindo a visualização de partidas e seus resultados, a visualização de equipes e campeonatos, a possibilidade de favoritá-los e a visualização de notícias atualizadas. Devido ao caráter acadêmico do projeto e à falta de orçamento para seu desenvolvimento, não foi possível utilizar APIs (MASSE, 2011) que atendessem aos requisitos necessários. Portanto, foi necessário desenvolver uma ferramenta de *Web Scraping* (ZHAO, 2017), que consiste em extrair dados de páginas da internet e salvá-los parcialmente ou em sua totalidade em um banco de dados.

No projeto, a cada intervalo de tempo pré-definido, a ferramenta captura dados de dois sites externos e extrai as informações necessárias. Um desses sites é o Placar de Futebol², de onde são extraídas informações sobre partidas, equipes e campeo-

¹ O site Arena Sport Club está disponível em: arenasportclub.site

² O site Placar de Futebol está disponível em: placardefutebol.com.br

natos. O segundo site é o FBref³, do qual são extraídos dados estatísticos variados, desde números sobre o ataque e a defesa de uma equipe até dados sobre o elenco, como idade média e salário semanal. Todas essas informações capturadas pela ferramenta são salvas no banco de dados do projeto e, posteriormente, consumidas pelo frontend das aplicações Web.

Utilizando o contexto do aplicativo, este projeto propõe a criação de uma ferramenta capaz de prever o resultado de uma partida de futebol. Com a variedade de dados disponíveis no banco de dados do projeto e o vasto conhecimento presente no ChatGPT, é possível combinar as tecnologias e as informações obtidas, com o objetivo de explorar diferentes utilizações desses dados e gerar numerosas possibilidades de resultados. Como parâmetro comparativo, utilizam-se os resultados reais das partidas e as probabilidades geradas pelas casas de apostas, que fazem uso de mecanismos e algoritmos avançados para gerar probabilidades de partidas, processos que costumam ser bastante complexos e custosos. O objetivo da utilização do ChatGPT neste contexto é avaliar sua capacidade de abstrair a complexidade envolvida nas previsões e de gerar resultados consistentes e satisfatórios. Além disso, busca-se analisar a viabilidade de seu uso em aplicações do mundo real, considerando não apenas a precisão dos resultados, mas também os custos associados à sua implementação.

A geração dos *prompts* (WHITE et al., 2023) e distribuição dos dados é uma parte de extrema importância ao utilizar LLMs. Um *prompt* pode ser uma simples pergunta ou um conjunto de instruções elaboradas e estruturadas. O objetivo é garantir que o modelo compreenda claramente o que o usuário deseja, sendo que quanto mais específico e bem estruturado for o *prompt*, melhor será a resposta gerada pelo modelo.

Portanto, é necessário formular *prompts* bem estruturados que contêm os dados disponibilizados na aplicação e as instruções principais para geração da previsão do resultado de uma partida, esses *prompts* serão enviados ao ChatGPT, que irá processar todas as informações e gerar a previsão. O objetivo final é possuir resultados satisfatórios, gerados em diferentes abordagens e que possam ser utilizados em um estudo comparativo para identificar a melhor solução. Será que o ChatGPT é capaz de processar uma grande quantidade de informações, advindas de múltiplas fontes e ter a capacidade de gerar resultados satisfatórios?

Na [Capítulo 2](#), será apresentada a fundamentação teórica do projeto, que abrange os conceitos e técnicas associadas às previsões de resultados de futebol e ao uso de inteligência artificial. Na [Capítulo 3](#), será explicada a metodologia utilizada, detalhando as etapas de desenvolvimento, implementação e avaliação das diferentes abordagens. Na [Capítulo 4](#), serão discutidos os resultados dos experimentos práticos, incluindo a comparação de desempenho entre as diversas estratégias de geração de *prompts*. Fi-

³ O site FBref está disponível em: fbref.com

nalmente, na [Capítulo 5](#), a conclusão que sintetiza os resultados do projeto e sugere direções futuras para aprimoramento e expansão da funcionalidade de previsão de resultados de partidas e o uso do ChatGPT em aplicações.

2 Fundamentação teórica

A evolução das tecnologias, especialmente no campo da inteligência artificial (IA), trouxe avanços significativos na maneira como analisamos e interpretamos grandes quantidades de dados. Entre essas inovações estão os Modelos de Linguagem de Larga Escala (LLMs) (BOMMASANI et al., 2022), que têm a capacidade de processar e gerar linguagem natural de forma avançada. Um LLM, como o ChatGPT, é treinado com uma vasta quantidade de dados textuais para aprender padrões e estruturas da linguagem, permitindo que ele entenda e gere respostas relevantes a uma ampla variedade de perguntas.

2.1 Evolução dos LLMs e surgimento do ChatGPT

A evolução dos LLMs se deu através de uma série de avanços significativos no campo do processamento de linguagem natural (NLP). Um dos marcos pioneiros foi o trabalho (VASWANI, 2017), que introduziu o modelo *Transformer*. Esse modelo revolucionou o processamento de linguagem ao eliminar a dependência de redes neurais recorrentes e convolucionais, utilizando uma arquitetura baseada exclusivamente em mecanismos de atenção. O *Transformer* permitiu que o processamento de texto ocorresse de forma mais eficiente e em paralelo, tornando possível treinar modelos em grandes quantidades de dados textuais de maneira mais escalável.

Logo após o *Transformer*, surgiram modelos que expandiram sua capacidade. Em 2018, o BERT (*Bidirectional Encoder Representations from Transformers*) foi lançado. Propondo uma abordagem inovadora ao processar o texto de maneira bidirecional, ou seja, analisando simultaneamente o contexto à esquerda e à direita de cada palavra (KENTON; TOUTANOVA, 2019). Isso trouxe um avanço considerável na compreensão do significado contextual das palavras, elevando o desempenho em várias tarefas de NLP, como tradução, resposta a perguntas e análise de sentimento.

Em paralelo ao BERT, o GPT (*Generative Pre-trained Transformer*) foi desenvolvido pela OpenAI. O GPT focou em uma abordagem não supervisionada para o treinamento em larga escala, utilizando grandes quantidades de dados textuais para predição de palavras subsequentes. O modelo se mostrou particularmente eficaz em tarefas de geração de texto, marcando o início da era dos LLMs geradores (RADFORD, 2018).

A partir dessas bases, a evolução dos LLMs continuou com o lançamento do GPT-2 (2019) e do GPT-3 (2020). O GPT-2 aumentou significativamente o número

de parâmetros do modelo em comparação com seu predecessor, permitindo maior fluência na geração de texto e em tarefas de compreensão de contexto. Já o GPT-3, com seus impressionantes 175 bilhões de parâmetros, ampliou ainda mais a capacidade do modelo em gerar textos quase indistinguíveis dos produzidos por humanos, consolidando os LLMs como ferramentas poderosas para diversas aplicações, como chatbots, geração de conteúdo e resumo automático de textos (BROWN et al., 2020).

Por fim, o modelo T5 (*Text-to-Text Transfer Transformer*) (XUE, 2020) levou os conceitos dos LLMs para um novo patamar, ao adotar uma abordagem unificada de entrada e saída em texto. Em vez de diferentes modelos para diferentes tarefas, o T5 transformou todos os problemas de NLP em tarefas de tradução de texto para texto, promovendo uma maior flexibilidade e simplificação no treinamento e na utilização dos modelos para diferentes aplicações de NLP.

2.2 Prompt Engineering

Nos artigos (WHITE et al., 2023) e (WHITE et al., 2024), são apresentados padrões de *prompts* que podem ser utilizados para direcionar o ChatGPT ao formato de resposta desejado ou para que ele assuma uma personalidade que ele possui conhecimento, fazendo com que ele seja direcionado ao objetivo do usuário com mais facilidade. Também são apresentadas categorizações para cada padrão de *prompt* e suas respectivas explicações. Em (WHITE et al., 2024), ocorreu um estudo mais específico sobre a utilização desses *prompts* para melhorar a qualidade de código, refatoração e outros aspectos do desenvolvimento de software.

Já no artigo (BROWN et al., 2020), ele explora os conceitos de *zero-shot*, *one-shot* e *few-shots learning*. Esses três conceitos são técnicas que podem ser utilizadas ao interagir com o ChatGPT. No *zero-shot learning*, o modelo é treinado para realizar uma tarefa sem exemplos de treinamento específicos para essa tarefa. No *one-shot learning*, apenas um único exemplo é fornecido para a tarefa. No *few-shots learning*, um pequeno número de exemplos é fornecido para a tarefa. Fornecer exemplos ao modelo de linguagem pode facilitar a compreensão do que está sendo solicitado e o que e como precisa ser respondido.

Neste projeto em específico, o foco será em formular *prompts* que facilitem ao LLM compreender as variáveis presentes que envolvem uma partida de futebol, e em seguida, utilizar esse conhecimento para realizar as previsões requeridas. Por ser uma atividade com muitos detalhes, será necessário utilizar algumas dessas técnicas apresentadas nos artigos citados acima, para que a ferramenta consiga abstrair as informações necessárias e gerar resultados satisfatórios.

2.3 Restrições do ChatGPT

Apesar de todos os potenciais benefícios que o ChatGPT e seu modelo podem oferecer, já foi observado em alguns estudos que ele exibe inconsistências e questões que demandam investigações mais aprofundadas. Em (BANG et al., 2023), foi constatado que o modelo pode ter quedas de desempenho ao interagir com múltiplas tarefas simultâneas. O que pode ser um problema para este trabalho, já que para processar todas as informações presentes em uma partida de futebol, é necessário manipular uma grande quantidade de dados e para isso é necessário possuir um grande poder de processamento e conhecimento. Será necessário observar esses aspectos, com o intuito de reportar possíveis quedas de rendimento.

Já em (CHEN; ZAHARIA; ZOU, 2023), é destacado que não é de conhecimento público, o que acontece com o modelo após cada atualização do ChatGPT. Também é apresentada uma comparação entre o mesmo modelo em duas datas distintas, em certos casos o ChatGPT perde precisão, e em outros ele tem um aumento de desempenho. Devido a esta falta de informação e consistência, a utilização do modelo de linguagem em larga escala pode perder sua viabilidade, devido às mudanças inesperadas e as dificuldades em identificar inconsistências nessas alterações. Para este trabalho, será necessário analisar os resultados obtidos entre os modelos disponibilizados, para verificar se a utilização do ChatGPT para o desenvolvimento de novas funcionalidades é viável ou não.

2.4 Aplicação de IA no futebol

O projeto Arena Sport Club foi concebido com o objetivo de criar uma funcionalidade capaz de prever os resultados de partidas de futebol, explorando o potencial do ChatGPT. A proposta envolve medir a capacidade desse modelo de linguagem em lidar com diferentes níveis de complexidade ao realizar previsões, avaliando se ele pode ser uma solução viável para o desenvolvimento de funcionalidades aplicáveis em contextos do mundo real.

Prever o resultado de uma partida de futebol envolve a análise de um conjunto complexo de variáveis que influenciam diretamente o desempenho das equipes e, conseqüentemente, o desfecho do jogo. Entre essas variáveis estão fatores estatísticos, como o número de gols marcados e sofridos, posse de bola, eficiência de finalizações e desempenho defensivo (PRASETIO; HARLILI, 2016). Além disso, fatores contextuais, como o local da partida (jogar em casa ou fora), condições climáticas, estado físico e emocional dos jogadores, lesões, suspensões e o histórico de confrontos entre as equipes também são considerados (RODRIGUES; PINTO, 2022). Métodos tradicionais de previsão, como regressão logística e modelos de Poisson, utilizam essas

variáveis para calcular a probabilidade de vitória, empate ou derrota de uma equipe, estabelecendo relações estatísticas entre elas (MAHER, 1982).

O uso do ChatGPT oferece algumas vantagens importantes nesse contexto. Sua capacidade de processar não apenas dados numéricos, mas também informações contextuais e linguísticas, permite que ele considere variáveis que vão além das estatísticas tradicionais. Além disso, o ChatGPT pode integrar diferentes fontes de dados, cruzando informações de maneira rápida e eficaz, algo que seria muito mais trabalhoso em modelos tradicionais baseados apenas em dados estruturados (VASWANI, 2017).

Entretanto, à medida que a quantidade de variáveis aumenta, também surgem desafios. O treinamento de um modelo como o ChatGPT envolve grandes quantidades de dados, e a adição de mais variáveis à previsão pode tornar o processo mais custoso em termos computacionais (BOMMASANI et al., 2022). Outro aspecto a ser considerado é que, embora o ChatGPT tenha um conhecimento vasto, ele depende da qualidade e da atualidade dos dados nos quais foi treinado. Portanto, ao adicionar mais variáveis ou tentar prever eventos muito específicos, há uma limitação natural relacionada à capacidade do modelo de lidar com todos esses dados sem perder precisão.

Dessa forma, o uso do ChatGPT no projeto Arena Sport Club visa encontrar um equilíbrio entre a simplicidade e a complexidade, utilizando o modelo para prever partidas com base em variáveis relevantes, mas também levando em consideração suas limitações quando confrontado com uma quantidade crescente de dados.

2.5 Trabalhos relacionados

No trabalho presente em (LIU et al., 2023) é explorado o uso de IA generativa na ciência de materiais, destacando como modelos baseados em aprendizado profundo podem revolucionar a descoberta e o design de novos materiais. Ao fornecer previsões precisas e eficientes sobre propriedades materiais e simulações, a IA generativa possibilita otimizações que anteriormente demandavam longos períodos de experimentação. Essa abordagem sugere que, assim como na ciência de materiais, o uso de LLMs na previsão esportiva pode transformar a forma como se analisam e se preveem os resultados de partidas futuras.

A previsão de resultados de futebol tem se tornado um campo de estudo cada vez mais relevante, impulsionado pelo avanço das técnicas de análise de dados e aprendizado de máquina. Diversos métodos e modelos têm sido propostos para abordar essa problemática, cada um com suas particularidades e benefícios.

Um exemplo notável é o uso de indicadores de rede de passes como ferra-

mentas preditivas para os resultados das partidas de futebol. No trabalho de (IEVOLI; PALAZZO; RAGOZINI, 2021), a análise das redes de passes permite capturar a dinâmica coletiva e a estratégia de jogo das equipes, oferecendo *insights* valiosos sobre a probabilidade de vitória, empate ou derrota. Esse método tem a vantagem de analisar a interação entre os jogadores em campo, refletindo a coesão da equipe e seu desempenho global. Os indicadores da rede de passes se mostraram eficazes na previsão de resultados em contextos de equipes que valorizam o jogo coletivo.

Outro método amplamente utilizado é a regressão logística, discutida no estudo de (PRASETIO; HARLILI, 2016). A regressão logística é um modelo estatístico que estima a probabilidade de ocorrência de um evento – neste caso, o resultado de uma partida de futebol – com base em variáveis explicativas, como estatísticas de ataque e defesa. Essa abordagem é bastante eficaz em prever resultados binários, como vitória ou derrota, utilizando uma função logística para classificar as probabilidades.

No campo do aprendizado de máquina, o trabalho de (RODRIGUES; PINTO, 2022) explora diferentes algoritmos para a previsão de resultados de futebol. O estudo comparou árvores de decisão, redes neurais e máquinas de vetores de suporte, destacando a capacidade dessas técnicas de identificar padrões complexos em grandes volumes de dados, como desempenho de jogadores, estatísticas de jogos anteriores e variáveis contextuais. O aprendizado de máquina, quando treinado com dados históricos e estatísticos, pode aumentar a precisão das previsões, uma vez que ajusta o modelo para reconhecer nuances que escapariam a abordagens mais simples.

Uma abordagem inovadora foi apresentada por (SHIN; GASPARYAN, 2014), que utilizou dados de videogames para prever partidas reais. No artigo, os autores argumentam que a indústria de videogames, ao tentar simular uma experiência realista no videogame FIFA 2015, acaba coletando dados muito precisos, como a movimentação dos jogadores, passes e chutes, que podem ser usados para resolver problemas do mundo real, como a previsão de resultados de futebol. Esse estudo sugere que, devido à similaridade entre o jogo simulado e o futebol real, os dados gerados em jogos como o FIFA podem ser aplicados em modelos preditivos para melhorar a precisão das previsões em ambientes reais.

O grande diferencial para este trabalho fica na parte da geração de *prompts* otimizados. Todo o trabalho relacionado a geração de algoritmos ou cálculos complexos será repassado ao ChatGPT, ele apenas receberá os *prompts* formulados junto aos dados filtrados e selecionados, para que ele compreenda o objetivo com mais facilidade. Todo o resto do processo será de obrigação dele pensar, processar e gerar o resultado. Se os resultados obtidos não atingirem um padrão satisfatório, é possível considerar a realocação de parte desse trabalho para os *prompts*, ou então simplificar o grau de complexidade da previsão. Isso pode ser alcançado por meio da exclusão de

algumas variáveis e dados, com o intuito de identificar um ponto em que os resultados se mostrem satisfatórios.

Após explorar as bases teóricas que delimitam o campo de estudo e estabelecer um entendimento crítico sobre os conceitos abordados, este estudo avança agora para a fase de metodologia. Na próxima seção, será detalhado toda a fase de planejamento do estudo, desde a apresentação da aplicação que servirá como base para a ferramenta de previsão, quanto a escolha do modelo de linguagem, a forma como os dados foram obtidos e como a ferramenta de previsão foi construída e como ela funciona.

3 Metodologia

No capítulo anterior, foram explorados os principais conceitos e o contexto teórico que envolvem este trabalho, revisando literaturas pertinentes que dialogam com o foco da pesquisa. Nesta seção, serão detalhados os procedimentos, técnicas e ferramentas empregadas na coleta e análise dos dados. Além disso, será apresentada a construção e o desenvolvimento da ferramenta de previsão para resultados de partidas de futebol, explicando como esses métodos se alinham com os objetivos propostos pelo estudo.

A ferramenta de previsão de partidas foi idealizada e desenvolvida como parte de um experimento científico conduzido exclusivamente por mim, com o objetivo de avaliar a eficácia de modelos de linguagem em prever resultados de partidas de futebol. O projeto foi estruturado com base em metodologias experimentais, onde foram definidos parâmetros específicos para a coleta de dados históricos, análise estatística e interpretação dos resultados. Utilizando o modelo de linguagem ChatGPT, foram implementadas diversas abordagens de engenharia de *prompt* para refinar a precisão das previsões. O desenvolvimento envolveu a subdivisão da ferramenta em diversas categorias personalizáveis, com o objetivo de manipular o formato, estrutura e os dados enviados, criando assim um ambiente interativo e controlado para testar hipóteses relacionadas à acurácia das previsões. Esse processo, ao longo de sua execução, permitiu a adaptação e validação do modelo, garantindo que os resultados fossem mensuráveis e replicáveis dentro de um contexto científico rigoroso.

Para o desenvolvimento e implementação das previsões de partidas, utilizou-se uma aplicação web denominada Arena Sport Club¹, criada durante a disciplina de Desenvolvimento de Aplicações Web, do curso de Ciência da Computação. Essa aplicação foi desenvolvida com o objetivo de fornecer uma plataforma acessível e interativa para a visualização de resultados e notícias relacionadas ao futebol. Através desta plataforma, os usuários podem visualizar os resultados de partidas de futebol, acessar resumos e estatísticas, além de obter detalhes e calendários sobre equipes e campeonatos.

A aplicação foi utilizada como ponto de partida para a idealização e construção da ferramenta de previsão proposta nesta pesquisa. As funcionalidades e limitações identificadas na aplicação original orientaram o desenvolvimento da solução da ferramenta de previsão de partidas, uma vez que os aspectos já consolidados foram aproveitados, enquanto as lacunas observadas serviram como referência para aprimorar

¹ Mais informações sobre o desenvolvimento e artefatos do projeto Arena Sport Club estão disponíveis em: github.com/ThiagoLuizBS/ArenaSportsClub

a nova ferramenta. Dessa forma, a pesquisa não apenas aprimora o escopo da aplicação original, mas também contribui com novos elementos tecnológicos voltados à previsão de resultados no contexto do futebol.

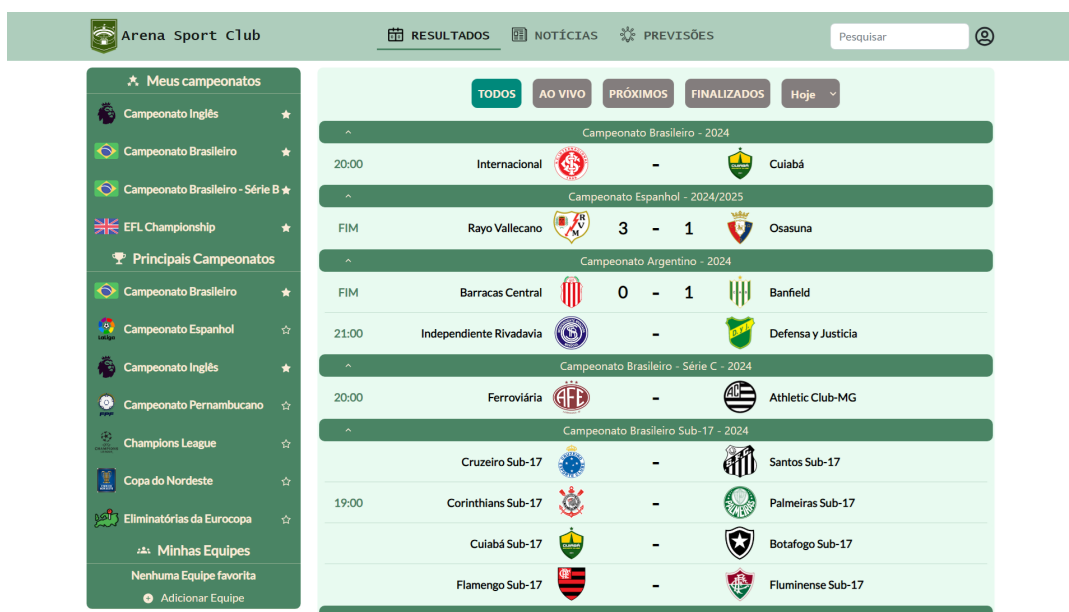
A implementação da aplicação foi realizada utilizando as tecnologias React e JavaScript no frontend, enquanto o backend foi desenvolvido com Express e Node.js. Para a provisão de dados e armazenamento de informações dos usuários, optou-se pelo uso do banco de dados não relacional MongoDB. A aplicação ficou em torno de 6 meses em período de desenvolvimento. A [Figura 1](#) apresenta a tela inicial da aplicação Arena Sport Club.

3.1 Seleção do LLM

Para a realização desta pesquisa, foi necessário selecionar o modelo de linguagem (LLM) mais adequado para a implementação das previsões de partidas. Diversos modelos foram analisados, incluindo o ChatGPT, o LLaMA (TOUVRON et al., 2023) e o PaLM (ANIL et al., 2023), entre outros. A escolha do modelo levou em consideração critérios fundamentais como a curva de aprendizado e a usabilidade. Devido às limitações de tempo do projeto, era essencial optar por um modelo que fosse de fácil adoção, permitindo que o desenvolvimento evoluísse rapidamente devido a fácil familiarização com suas funcionalidades.

Outro aspecto decisivo foi a disponibilidade e a longevidade do modelo, garantindo que ele estivesse operacional durante todo o período de pesquisa. A continuidade

Figura 1 – Tela inicial do Arena Sport Club



Fonte: Autor (2024)

do funcionamento do modelo após a conclusão do estudo também foi considerada, para que os resultados pudessem ser reproduzidos e validados futuramente.

Diante da análise comparativa entre os modelos, o ChatGPT demonstrou superioridade em relação ao PaLM e ao LLaMA em vários aspectos. O PaLM, apesar de ser um modelo robusto e avançado, apresentou uma curva de aprendizado mais elevada e menos acessível em termos de documentação e usabilidade, o que dificultaria sua implementação no curto prazo da pesquisa. Já o LLaMA, embora promissor, ainda é relativamente novo no mercado e apresenta incertezas quanto à sua continuidade e suporte a longo prazo.

O ChatGPT se destacou por combinar facilidade de uso, vasta documentação e atualizações frequentes, o que o tornou a opção mais segura e eficiente para atender aos requisitos do projeto. Esses fatores, aliados à sua popularidade e estabilidade, foram decisivos para a escolha, garantindo uma implementação mais rápida e confiável.

Após a seleção do modelo, o próximo passo foi determinar a forma de utilização do ChatGPT. Uma das opções consideradas foi o uso de sua API, que permite o envio de *prompts* para um endpoint, retornando as respostas de maneira semelhante à interação realizada via navegador ou aplicativo. Outra alternativa foi a utilização do modelo de *Fine-tuning* do ChatGPT, que possibilita a personalização do modelo com uma base de dados própria para treinamento. De maneira simplificada, o processo de *Fine-tuning* envolve a preparação e o upload dos dados de treinamento, seguido pela fase de treinamento do modelo por um determinado período. Após a conclusão do treinamento, é possível avaliar os resultados obtidos. Caso os resultados sejam satisfatórios, o modelo treinado pode ser salvo e utilizado de forma semelhante à API convencional. Além disso, o treinamento pode ser retomado posteriormente para novas iterações ou aprimoramentos.

Diante das opções apresentadas, foi necessário analisar os pontos positivos e negativos de cada abordagem, a fim de identificar a melhor solução considerando as limitações do projeto. O uso da API de forma tradicional oferece vantagens como facilidade de uso e uma documentação abrangente. A interação com a API é direta: basta preencher os parâmetros necessários, elaborar os *prompts* e enviar a requisição. No entanto, a API apresenta algumas limitações, sendo a principal delas a incapacidade de armazenar o histórico das conversas realizadas. Cada requisição é tratada como uma nova interação, retornando apenas uma única resposta. Embora essa limitação possa ser contornada ao incluir o histórico da conversa no *prompt* da requisição, tal abordagem encontra restrições. A API impõe limites de *tokens* por requisição, e cada *token* enviado ou recebido tem um custo associado. Conseqüentemente, à medida que o histórico da conversa se expande, o custo das requisições aumenta. Considerando as restrições financeiras do projeto, essa solução de inclusão do histórico torna-se

inviável a longo prazo.

A análise da opção de *Fine-tuning* revela diversos aspectos positivos, incluindo uma maior facilidade do modelo em compreender as requisições e uma melhora na qualidade das respostas geradas. A personalização do treinamento permite a criação de um modelo adaptado a contextos ou situações específicas, o que pode resultar em uma solução altamente eficiente. No entanto, o *Fine-tuning* também apresenta desafios significativos. Quanto maior o volume de dados de treinamento, maior será o custo associado. Para obter um modelo preciso e alinhado com as características desejadas, seria necessária uma vasta base de dados para o treinamento. No contexto deste projeto, isso implicaria a preparação de uma quantidade substancial de dados históricos de partidas e seus resultados. Além do custo do treinamento do modelo, haveria também despesas contínuas com as requisições e respostas geradas pelo modelo personalizado. Considerando o tempo reduzido e a limitação orçamentária do projeto, a opção de *Fine-tuning* torna-se inviável.

Dessa forma, a alternativa mais adequada às condições da pesquisa é o uso da API de forma tradicional, com o objetivo de minimizar os custos o máximo possível. Considerou-se a base de dados dos modelos do ChatGPT como potencialmente suficiente para gerar respostas satisfatórias no contexto deste projeto. No entanto, a eficácia dessa abordagem pôde ser plenamente avaliada apenas após a realização dos testes, quando os resultados concretos foram analisados para verificar se atendiam aos requisitos estabelecidos.

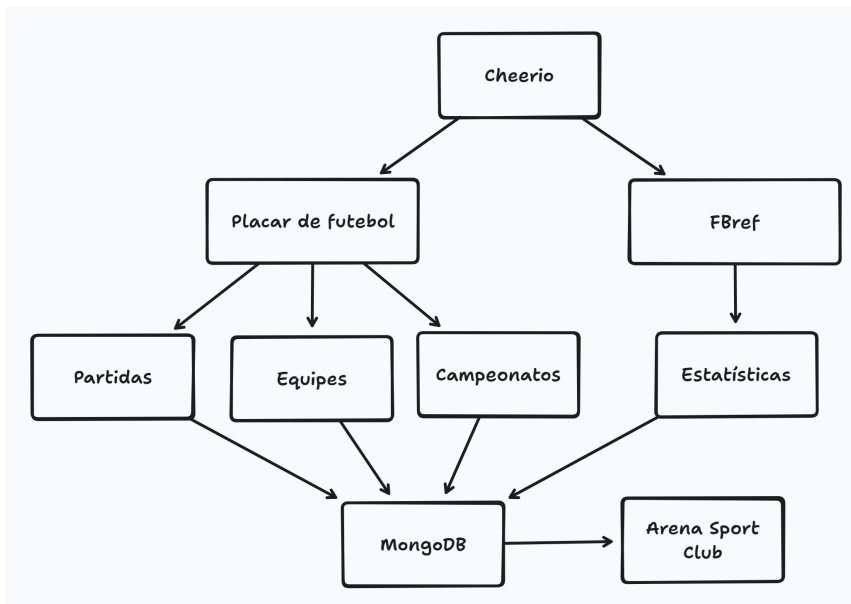
3.2 Obtenção dos dados

Para que as funcionalidades da aplicação pudessem operar corretamente, foi necessário dispor de uma grande quantidade de dados. No contexto do futebol, existem diversas APIs disponíveis no mercado, entretanto, a maioria delas são pagas. As APIs que oferecem planos gratuitos não atendem plenamente às necessidades da aplicação, que requer a disponibilização de informações sobre diversas partidas e campeonatos. Dado que o projeto não contava com orçamento durante sua implementação, a alternativa mais viável escolhida foi a utilização de *web scrapers* para a obtenção dos dados necessários.

Para automatizar a extração desses dados, o *web scraper* foi desenvolvido utilizando a biblioteca Cheerio. Essa ferramenta, ao receber uma URL, captura todo o HTML da página, permitindo a manipulação dos dados por meio da busca por elementos que se repetem e que compartilham a mesma classe ou estrutura HTML. Para garantir que a raspagem fosse realizada de maneira eficiente e contínua, foi criado um intervalo pré-definido para cada parte do *web scraper*, permitindo que a cada cinco

minutos uma etapa específica do processo fosse executada separadamente. Isso assegurou a coleta regular e sistemática das informações necessárias para alimentar o banco de dados da aplicação. No Fluxograma 1, é apresentado um esboço da estrutura dos *web scrapers* utilizados na aplicação.

Fluxograma 1: Estrutura dos web scrapers utilizados no Arena Sport Club



Fonte: Autor (2024)

A raspagem de dados foi utilizada em dois cenários distintos. No primeiro cenário, para atender às necessidades básicas da aplicação, como a obtenção de informações sobre partidas, campeonatos e equipes, foi realizada a raspagem de dados no site Placar de Futebol. Esse processo foi dividido em quatro etapas. Na 1ª etapa, a raspagem ocorre na tela inicial do site, onde são listadas todas as partidas do dia. Cada uma dessas partidas é percorrida, e algumas informações-chave são armazenadas. As três principais informações extraídas são: o link para a página da partida, o link para a página do campeonato correspondente, e os links para as páginas das duas equipes que estão se enfrentando. A Figura 2 ilustra as partidas disponíveis na tela inicial do site.

Nas etapas seguintes, os três links armazenados são utilizados para realizar novas raspagens, permitindo a extração de todas as informações necessárias para a construção da API. A 2ª etapa ocorre na página de cada partida, são coletados dados como detalhes do jogo, estatísticas, escalações e um resumo dos acontecimentos. A 3ª etapa acontece na página da equipe, são extraídas informações sobre a equipe, títulos conquistados, últimos e próximos jogos. Por fim, a 4ª etapa é realizada na página do campeonato, onde são armazenadas as estatísticas, a classificação e os resultados das partidas. Com essas informações armazenadas separadamente, torna-se possível

Figura 2 – Tela inicial do site Placar de Futebol



Fonte: Autor (2024)

replicar a estrutura do site original na aplicação Arena Sport Club.

Embora as informações já coletadas fossem suficientes para gerar previsões de resultados das partidas, a adição de mais dados ao LLM poderia aumentar a confiabilidade e consistência dos resultados finais. Com isso em mente, também foi realizada a raspagem de dados em outro site, o FBref. Em comparação ao Placar de Futebol, o FBref oferece uma quantidade significativamente maior de informações, porém, as políticas de uso do site limitam bastante a possibilidade da utilização de *web scraping* de forma mais abrangente. Caso sejam realizadas mais de 10 solicitações por minuto ao site pelo mesmo IP, esse IP receberá um bloqueio progressivo que o impede de acessar o site. Portanto foi necessário selecionar cuidadosamente o conteúdo que seria capturado pelo *web scraper*.

Conforme ilustrado na Figura 3, o FBref disponibiliza uma vasta gama de dados estatísticos sobre campeonatos, com mais de 60 tipos de estatísticas organizadas em diferentes categorias. Visando utilizar alguns desses dados nas previsões de partidas, foi realizada a raspagem de dados de páginas de campeonatos previamente selecionados, como já tinha sido apresentado no Fluxograma 1. Os dados extraídos foram então armazenados junto com os dados dos campeonatos obtidos do Placar de Futebol e podem ser utilizados em conjunto tanto para a apresentação das informações dos campeonatos no site quanto para a previsão dos resultados das partidas.

Durante o período de pesquisa deste trabalho, a utilização de *web scraping* como ferramenta para a extração de dados enfrentou alguns desafios. Em algumas ocasiões, devido a mudanças na estrutura do HTML ou atualizações no layout de

Figura 3 – Tela de estatísticas do campeonato inglês 2023/2024 no site FBref

The screenshot displays two tables from the FBref website. The top table is 'Squad Standard Stats' for the 2023-2024 Premier League, listing 20 teams with various performance metrics. The bottom table is 'Squad Goalkeeping' for the same league, listing the same 20 teams with goalkeeping-specific statistics.

Squad		#	PI	Age	Pos	Playing Time		Performance										Expected			Progression		Per 90 Minutes									
Squad	#	PI	Age	Pos	MP	Starts	Min	90s	Gls	Ass	G+A	G-PK	PK	PKatt	CrdY	CrdR	xG	nxpG	xAG	nxpG+xAG	PrgC	PrgP	Gls	Ass	G+A	G-PK	G+A-PK	xG	xAG	xG+xAG	nxpG	nxpG+xAG
Arsenal	25	25.0	58.2	38	418	3,420	38.0	86	62	148	76	10	10	64	2	76.1	68.5	55.3	123.8	821	2106	2.26	1.63	3.89	2.00	3.63	2.00	1.46	3.46	1.80	3.26	
Aston Villa	31	26.9	53.1	38	418	3,420	38.0	72	55	127	68	4	4	94	2	63.3	60.3	48.6	108.9	775	1451	1.89	1.45	3.34	1.79	3.24	1.67	1.28	2.95	1.59	2.87	
Bournemouth	30	25.7	44.4	38	418	3,420	38.0	52	39	91	50	2	3	79	3	55.9	53.6	40.5	94.0	709	1314	1.37	1.03	2.39	1.32	2.34	1.47	1.06	2.54	1.41	2.47	
Brentford	28	26.6	45.2	38	418	3,420	38.0	54	38	92	51	3	3	89	2	58.2	56.0	42.7	98.8	505	1283	1.42	1.00	2.42	1.34	2.34	1.53	1.12	2.66	1.47	2.60	
Brighton	30	25.9	59.8	38	418	3,420	38.0	50	36	86	45	5	6	90	3	56.8	52.0	41.6	93.8	864	1780	1.32	0.95	2.26	1.18	2.13	1.49	1.09	2.59	1.37	2.47	
Burnley	31	24.5	47.4	38	418	3,420	38.0	40	30	70	37	3	3	77	7	40.6	38.3	29.7	68.0	727	1181	1.05	0.79	1.84	0.97	1.76	1.07	0.78	1.85	1.01	1.79	
Chelsea	32	24.2	58.6	38	418	3,420	38.0	76	51	127	65	11	12	109	4	74.5	65.1	51.4	116.5	873	1615	2.00	1.34	3.34	1.71	3.05	1.96	1.35	3.31	1.71	3.06	
Crystal Palace	26	27.7	42.3	38	418	3,420	38.0	56	43	99	52	4	4	73	2	48.6	45.4	38.7	84.1	506	1176	1.47	1.13	2.61	1.37	2.50	1.28	1.02	2.30	1.20	2.21	
Everton	26	27.1	40.8	38	418	3,420	38.0	40	17	57	38	2	3	82	1	54.0	51.6	38.1	89.8	495	1129	1.05	0.45	1.50	1.00	1.45	1.42	1.00	2.43	1.36	2.36	
Fulham	25	28.4	50.6	38	418	3,420	38.0	54	40	94	52	2	2	84	4	50.8	49.2	37.6	86.8	670	1467	1.42	1.05	2.47	1.37	2.42	1.34	0.99	2.33	1.30	2.29	
Liverpool	30	26.3	61.3	38	418	3,420	38.0	80	60	140	74	6	9	69	5	87.8	80.4	63.7	144.1	924	2110	2.11	1.58	3.68	1.95	3.53	2.31	1.68	3.98	2.12	3.79	
Luton Town	28	26.7	42.4	38	418	3,420	38.0	49	33	82	44	5	5	71	0	42.4	38.6	29.5	68.2	675	1113	1.29	0.87	2.16	1.16	2.03	1.12	0.78	1.89	1.02	1.79	
Manchester City	25	26.6	65.2	38	418	3,420	38.0	94	69	163	85	9	10	53	2	80.5	72.9	60.2	133.1	1127	2060	2.47	1.82	4.29	2.24	4.05	2.12	1.58	3.70	1.92	3.50	
Manchester United	31	26.1	50.6	38	418	3,420	38.0	57	36	93	51	6	7	83	1	56.5	51.0	40.1	91.0	759	1479	1.50	0.95	2.45	1.34	2.29	1.49	1.06	2.54	1.34	2.40	
Newcastle Utd	32	27.2	52.2	38	418	3,420	38.0	83	55	138	75	8	9	79	1	76.0	69.0	51.3	120.3	737	1557	2.18	1.45	3.63	1.97	3.42	2.00	1.35	3.35	1.82	3.17	
Nottingham Forest	33	25.7	40.6	38	418	3,420	38.0	49	43	92	48	1	1	82	3	49.9	49.1	38.0	87.1	604	1140	1.29	1.13	2.42	1.26	2.39	1.31	1.00	2.31	1.29	2.29	
Sheffield Utd	35	25.7	35.8	38	418	3,420	38.0	31	18	49	26	5	5	101	5	38.3	34.4	25.8	60.2	343	884	0.82	0.47	1.29	0.68	1.16	1.01	0.68	1.69	0.91	1.58	
Tottenham	29	25.2	61.6	38	418	3,420	38.0	69	58	127	67	2	2	92	4	68.2	66.7	57.4	124.1	957	2107	1.82	1.53	3.34	1.76	3.29	1.79	1.51	3.31	1.75	3.27	
West Ham	26	28.3	41.1	38	418	3,420	38.0	58	40	98	54	4	5	82	3	52.3	48.4	34.8	83.2	556	1207	1.53	1.05	2.58	1.42	2.47	1.38	0.92	2.29	1.27	2.19	
Wolves	27	26.6	48.9	38	418	3,420	38.0	47	36	83	43	4	4	102	4	46.7	43.6	31.8	75.4	671	1180	1.24	0.95	2.18	1.13	2.08	1.23	0.84	2.07	1.15	1.98	

Squad		#	MP	Starts	Min	90s	Performance										Penalty Kicks			
Squad	#	MP	Starts	Min	90s	GA	GA90	SoTA	Saves	Save%	W	D	L	CS	CS%	PKatt	PKA	PKsv	PKm	Save%
Arsenal	2	38	38	3,420	38.0	29	0.76	83	54	67.5	28	5	5	18	47.4	3	2	1	0	33.3

Fonte: Autor (2024)

ambos os sites apresentados, o código desenvolvido para automatizar a coleta de informações necessitava de ajustes. Tais modificações, tornavam partes do código de *scraping* obsoletas, resultando em falhas na extração dos dados. Esse processo exigia uma constante revisão e atualização do código para garantir a continuidade e a precisão na coleta das informações, consumindo tempo adicional e recursos durante a pesquisa. Essas dificuldades evidenciam a natureza dinâmica dos sites na web e a necessidade de manter uma abordagem flexível e adaptável ao utilizar técnicas de *web scraping*.

3.3 Funcionalidade de previsão do resultado de partidas

Com o LLM escolhido e a forma de sua utilização definida, juntamente com os dados extraídos e armazenados, o próximo passo para gerar as previsões dos resultados das partidas é construir e estruturar todo o sistema que integrará os dados necessários, estruturá-los em um *prompt* e enviar a requisição para a API do ChatGPT. A Figura 4 ilustra como essa funcionalidade é apresentada na aplicação Arena Sport Club.

Primeiramente, o usuário deve selecionar a partida para a qual deseja realizar a previsão. Em seguida, ele deve selecionar as opções disponibilizadas em cada categoria presente na ferramenta, sendo que cada categoria possui uma finalidade específica. Há, por exemplo, a categoria do modelo a ser utilizado pelo ChatGPT, enquanto outras categorias correspondem a parâmetros exigidos pela API na requisição. Existem ainda

Figura 4 – Tela da funcionalidade de previsão de resultados de partidas no Arena Sport Club

Previsão de partidas

Selecione a partida e as informações que serão utilizadas para gerar a previsão

Selecionar uma partida

Modelo

- gpt-3.5-turbo
- gpt-4
- gpt-4o

Temperatura

- 0
- 1
- 2

Número de exemplos para o modelo

- Zero-shot
- One-shot
- Three-shots

Número de partidas no histórico

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Estatísticas das partidas

- Posse de bola (%)
- Total de passes
- Passes corretos (%)
- Total de chutes
- Chutes no gol
- Escanteios
- Faltas cometidas

Apresentação das estatísticas

- Por partida
- Por média

Estatísticas do campeonato

- Posição
- Pts/J/V/E/D
- GM/GS/SG
- xG/xGA/xGD
- Sh/SoT/SoT%
- Poss/Cmp/Cmpp
- PrgC/PrgP
- Faltas cometidas
- Cartões amarelos e vermelhos
- Clean Sheets
- Escanteios cobrados
- Idade média do elenco
- Salário do elenco

Histórico de partidas a serem analisadas

- Todas
- Apenas mandante ou visitante

Gerar previsão

Fonte: Autor (2024)

categorias que permitem definir a forma como o *prompt* será enviado e estruturado, e, por fim, categorias que determinam quais dados estatísticos serão adicionados aos *prompts*.

A integração da ferramenta no projeto Arena Sport Club é feita diretamente com o backend da aplicação. No backend é realizada a integração com a API do ChatGPT e toda a parte de estruturação e definição de quais dados devem ser enviados na requisição.

3.3.1 Categorização dos dados para previsão

Cada categoria do sistema de previsão de partidas desempenhou um papel essencial no desenvolvimento da ferramenta, contribuindo para a criação de um ambiente robusto, diversificado e funcional. Para garantir a eficácia do sistema, foi realizado um benchmarking detalhado, no qual foram analisadas diversas abordagens nos estudos relacionados analisados. Esse processo envolveu a comparação de diferentes técnicas e parâmetros utilizados por outros sistemas de previsão, permitindo identificar as melhores práticas e ajustar o modelo às necessidades específicas do projeto.

A personalização da ferramenta foi guiada pela necessidade de equilibrar a precisão e a flexibilidade, garantindo que as previsões geradas fossem adaptáveis a diferentes cenários. Essa customização envolveu a otimização de *prompts* e a implementação de ajustes contínuos, que foram testados e validados durante o processo experimental, permitindo que a ferramenta se tornasse altamente eficaz na previsão

dos resultados das partidas.

Na categoria de “**Modelo**”, é possível selecionar o modelo de linguagem que receberá a requisição, sendo que cada um deles apresenta características distintas. O **Quadro 1** apresenta o custo para cada 1 milhão de *tokens* em cada modelo, com os valores de entrada (input) e saída (output) cobrados separadamente. O modelo *GPT-3.5-turbo* é o mais antigo disponibilizado na tabela e é o modelo gratuito disponibilizado no website e aplicativo do ChatGPT, durante a época de realização da pesquisa. Apesar disso, ele demonstra uma boa capacidade de contextualização e resposta, sendo o modelo mais acessível em termos de custo. O modelo *GPT-4-turbo* é uma evolução de seu antecessor, destacando-se na execução de tarefas mais complexas, devido ao seu elevado nível de compreensão textual e linguística em comparação ao *GPT-3.5-turbo*. Além disso, possui uma base de dados de treinamento maior e mais atualizada; contudo, é também o modelo mais caro, sendo 20 vezes mais custoso que o *GPT-3.5-turbo*. Por fim, o modelo *GPT-4o* é uma versão atualizada do *GPT-4-turbo*. Ele é mais ágil na geração de respostas e possui uma base de treinamento mais recente, mas ainda apresenta menor eficiência em tarefas complexas quando comparado ao *GPT-4-turbo*. Seu custo é um meio termo entre o valor dos outros dois modelos apresentados.

Quadro 1: Custos de utilização da API do ChatGPT para cada modelo

Modelo	Valores para cada 1M de <i>tokens</i>
<i>GPT-4-turbo</i>	\$10,00 / <i>tokens</i> de pergunta \$30,00 / <i>tokens</i> de resposta
<i>GPT-4o</i>	\$5,00 / <i>tokens</i> de pergunta \$15,00 / <i>tokens</i> de resposta
<i>GPT-3.5-turbo</i>	\$0,50 / <i>tokens</i> de pergunta \$1,50 / <i>tokens</i> de resposta

Fonte: Autor (2024)

A próxima categoria refere-se a um parâmetro opcional da API, conhecido como “**Temperatura**”. A temperatura é um parâmetro que determina o grau de aleatoriedade ou criatividade do modelo, sendo representado por um valor decimal que pode variar de 0 a 2. Caso não seja especificado, o valor padrão é 1. No sistema de previsão, é possível escolher entre três opções: ao selecionar 0, o modelo funcionará de maneira mais focada e determinística possível; ao escolher o valor 2, o modelo adotará um comportamento mais aleatório e criativo em suas respostas; já a opção padrão representa um equilíbrio entre os dois extremos.

A categoria seguinte é chamada de “**Número de exemplos para o modelo**”, ela define o número de exemplos que serão adicionados aos *prompts*. Com a opção *Zero-shot*, nenhum exemplo de treinamento é enviado ao modelo. Na opção *One-shot*,

o modelo recebe um exemplo de pergunta e resposta. Na última opção, *Three-shots*, são fornecidos três exemplos ao modelo.

A categoria "**Número de partidas no histórico**" também influencia na estruturação do *prompt*, porém, a partir dessa categoria, os dados começam a ser manipulados. Cada opção define quantas partidas anteriores de cada equipe devem ser inseridas aos *prompts*, variando de 1 a 10 partidas.

A categoria subsequente permite selecionar quais dessas partidas históricas devem ser consideradas, ela é chamada de "**Histórico de partidas a serem analisadas**" e possui as seguintes opções: o usuário pode optar por incluir todas as partidas ou restringir a seleção apenas às partidas em que o time mandante jogou como mandante ou vice-versa, essa categoria foi criada após analisar alguns trabalhos (PRASETIO; HARLILI, 2016)(RODRIGUES; PINTO, 2022)(Football Data, 2003), que destacavam a separação dos dados de partidas mandantes ou visitantes.

As três últimas categorias estão relacionadas aos dados estatísticos das partidas e do campeonato. As "**Estatísticas das partidas**" subdividem-se em sete métricas, extraídas por meio de *web scraping* do site Placar de Futebol, elas são: posse de bola (%), total de passes, passes corretos (%), total de chutes, chutes no gol, escanteios e faltas cometidas.

As "**Estatísticas do campeonato**", referentes à partida selecionada, foram obtidas através do site FBref. Como o site oferece uma vasta gama de estatísticas, sendo mais de 60 tipos relacionados ao campeonato, foi necessário realizar uma filtragem criteriosa para selecionar as estatísticas mais relevantes para o estudo. A seleção baseou-se na frequência de uso e na relevância dessas estatísticas em análises esportivas, conforme identificado nos trabalhos relacionados (SHIN; GASPARYAN, 2014)(RODRIGUES; PINTO, 2022) e em práticas comuns de jornalismo esportivo.

Ao total foram selecionadas 27 estatísticas de campeonatos a serem disponibilizadas na ferramenta de previsão. Elas podem se subdividir em alguns grupos, o 1º grupo remete as estatísticas de classificação do campeonato, onde contém as informações da posição classificatória da equipe, pontos conquistados, jogos disputados, vitórias, empates e derrotas. O 2º grupo contém informações sobre a ofensividade da equipe, entre elas temos gols marcados, gols sofridos, saldo de gols, gols esperados (*xG*), gols esperados contra (*xGA*), gols esperados diferença (*xGD*), chutes, chutes ao alvo e porcentagem de chutes ao alvo. O 3º grupo destaca informações sobre os passes da equipe, contendo a média da posse de bola, passes completos, porcentagem de passes completos, corridas progressivas e passes progressivos. O 4º grupo possui estatísticas gerais que acontecem nas partidas, como cartões amarelos aplicados, cartões vermelhos, faltas cometidas, escanteios cobrados e jogos sem sofrer gols (*clean sheets*). Por fim, o 5º grupo contém informações sobre as equipes em si, como média

de idade e os custos de salários semanais do elenco.

Por fim, há uma categoria denominada "**Apresentação das estatísticas**". Cada opção nessa categoria altera a forma como as estatísticas da partida são estruturadas no *prompt*. Se a opção "por partida" for selecionada, cada partida terá sua descrição e estatísticas detalhadas no *prompt*. Se a opção "por média" for escolhida, cada partida terá sua descrição, mas as estatísticas serão apresentadas apenas uma vez, com os valores sendo a média de todas as partidas.

3.3.2 Elaboração dos prompts

A elaboração e estruturação de *prompts* para a requisição à API do ChatGPT constituem uma etapa essencial no processo de interação com o modelo. A criação de *prompts* eficientes é fundamental para a obtenção de respostas precisas e relevantes, podendo, assim, melhorar significativamente a qualidade das interações. Diversas técnicas e estratégias podem ser adotadas para atingir esse objetivo, cada uma com características e aplicações específicas.

A seguir, é apresentado o principal *prompt* que foi formulado, logo após na [subseção 3.3.2.1](#) será demonstrada as abordagens adotadas para garantir o sucesso na obtenção de previsões consistentes e alinhadas com os objetivos da pesquisa. Ao aplicar uma combinação dessas técnicas, foi possível otimizar a estruturação dos *prompts* para a previsão de resultados de partidas de futebol. Os *prompts*, desenvolvidos com base em metodologias específicas, foram configurados de maneira a maximizar a precisão e a relevância das respostas fornecidas pelo modelo. Os *prompts* foram testados e validados de forma incremental, com o objetivo principal do modelo compreender o que foi apresentado e ele gerar a previsão no formato solicitado.

Você é um especialista matemático e grande conhecedor de futebol. Irei te fornecer algumas informações sobre uma partida de futebol que vai ocorrer. Essas informações incluem o nome de cada equipe, o local da partida e o histórico de partidas de cada equipe. O histórico contém informações das últimas X partidas, incluindo resultado, estatísticas, data e local da realização da partida. Também haverá estatísticas do campeonato em que a partida será realizada. Seu trabalho será analisar todas essas informações em conjunto, especialmente o histórico de jogos de cada equipe e fazer uma previsão do resultado da partida que irá acontecer. Sua resposta deve seguir o modelo de tabela abaixo:

Time X x Time Y

Probabilidade Time X vitória: 45% | Empate: 20% | Probabilidade Time Y vitória: 25%

Agora irei te fornecer uma partida por vez, cada uma com suas informações e o histórico de partidas de cada equipe. Analise minuciosamente todas as informações disponibilizadas e após isso me forneça apenas a tabela como resposta.

3.3.2.1 Técnicas utilizadas nos prompts desenvolvidos

De acordo com (WHITE et al., 2023), a aplicação de padrões específicos de *prompt engineering*, como a utilização de cenários condicionais e a inclusão de exemplos, pode otimizar ainda mais a performance do modelo ao fornecer contexto e especificidade, resultando em respostas mais precisas e adequadas para a tarefa em questão. O artigo também discute outras técnicas que podem ser empregadas para melhorar a compreensão do modelo em relação à requisição, gerando respostas mais satisfatórias.

Um dos primeiros passos na criação de *prompts* eficazes é garantir que eles sejam claros e bem estruturados. A clareza envolve a eliminação de ambiguidades e a especificação precisa do que se espera do modelo. Fornecer contexto relevante no *prompt* auxilia o modelo de linguagem a compreender melhor a tarefa ou a pergunta. No *prompt* formulado, inicialmente é atribuído uma persona a ele, como "Você é um especialista matemático e grande conhecedor de futebol", isso ajuda o modelo a delimitar a área de interesse, gerando respostas mais centradas e focadas no contexto desejado.

Em seguida, é apresentada uma explicação sobre as informações que serão fornecidas ao modelo, detalhando a estrutura de algumas dessas informações. Após essa etapa, especifica-se claramente o que o modelo precisa realizar ao receber tais informações. Essa abordagem direta e bem específica geralmente resulta em melhores respostas, uma vez que a especificidade reduz o escopo das possíveis respostas, permitindo ao modelo concentrar-se no aspecto exato da questão.

Depois das explicações e detalhes, é indicado o formato desejado para as respostas, bem como as informações que devem ser incluídas. Além disso, é fornecido um exemplo de como a resposta deve ser estruturada, o que pode ser extremamente útil para guiar o modelo de linguagem. Os exemplos oferecem uma referência clara do tipo de resposta esperada, o que pode melhorar significativamente a precisão da saída do modelo.

Por fim, o *prompt* informa que será enviada apenas uma partida por vez e solicita que o modelo analise todas as informações minuciosamente. Essa instrução enfatiza a necessidade de o modelo não gerar a resposta de forma apressada, mas sim utilizar todas as informações disponíveis para produzir uma resposta no formato solicitado.

3.3.2.2 Utilização dos prompts na API do ChatGPT

Ao utilizar a API do ChatGPT, é possível simular uma conversa por meio do sistema de mensagens da API. Nesse contexto, são adicionados um ou mais *prompts* a uma lista, e cada *prompt* possui uma *role*. A *role* refere-se aos papéis que podem ser atribuídos às mensagens dentro de uma conversa, podendo ser do tipo sistema, usuário ou assistente.

O tipo sistema define o comportamento ou a personalidade do modelo ao longo de toda a conversa. O tipo usuário representa as mensagens enviadas pelo usuário, ou seja, o que o usuário deseja que o modelo responda ou processe. Já o tipo assistente reflete as respostas ou ações que o modelo deve tomar.

Utilizando esse sistema da API, o *prompt* apresentado anteriormente utiliza a *role* sistema. Na tela da funcionalidade de previsão de resultados de partidas no Arena Sport Club (Figura 4), o usuário pode selecionar, na categoria "número de exemplos para o modelo", as opções *Zero*, *One* ou *Three-shots learning*. Caso opte por uma das duas últimas opções, serão adicionadas à requisição da API mensagens com as *roles* usuário e assistente. Essas mensagens servem como exemplos de previsão de resultados de partidas, contendo o *prompt* solicitado pelo usuário e a resposta fornecida pelo modelo.

Com essa abordagem, o modelo pode adaptar-se a diferentes contextos e reduzir inconsistências, uma vez que os exemplos auxiliam na clareza do que o usuário deseja, diminuindo a probabilidade de mal-entendidos ou respostas fora do contexto.

3.4 Exemplo da funcionalidade de previsão do resultado de partidas

Após a apresentação detalhada de todas as etapas que compõem a ferramenta de previsões de partidas, é possível ilustrar um exemplo de sua utilização. Primeiramente, deve-se escolher uma partida para realizar a previsão. Para isso, basta acessar a tela principal da partida e clicar no botão "Prever partida". Em seguida, o usuário será redirecionado para a tela de previsão, conforme ilustrado na Figura 5.

Nessa etapa, é necessário selecionar as opções disponíveis em cada categoria, a fim de personalizar o *prompt*. Por fim, basta clicar em "Gerar previsão" e, em alguns segundos, o resultado da requisição será disponibilizado, conforme mostrado na Figura 6. O resultado que o modelo apresenta são as probabilidades de vitória de cada equipe e a probabilidade de empate da partida, ele também disponibiliza as partidas que foram adicionadas ao histórico do *prompt*.

É possível alterar as opções de cada categoria e gerar uma nova previsão; ao alternar as opções, diferentes resultados podem ser obtidos. Para avaliar a precisão e

Figura 5 – Tela da funcionalidade de previsão de resultados de partidas no Arena Sport Club, após selecionar uma partida para gerar a previsão

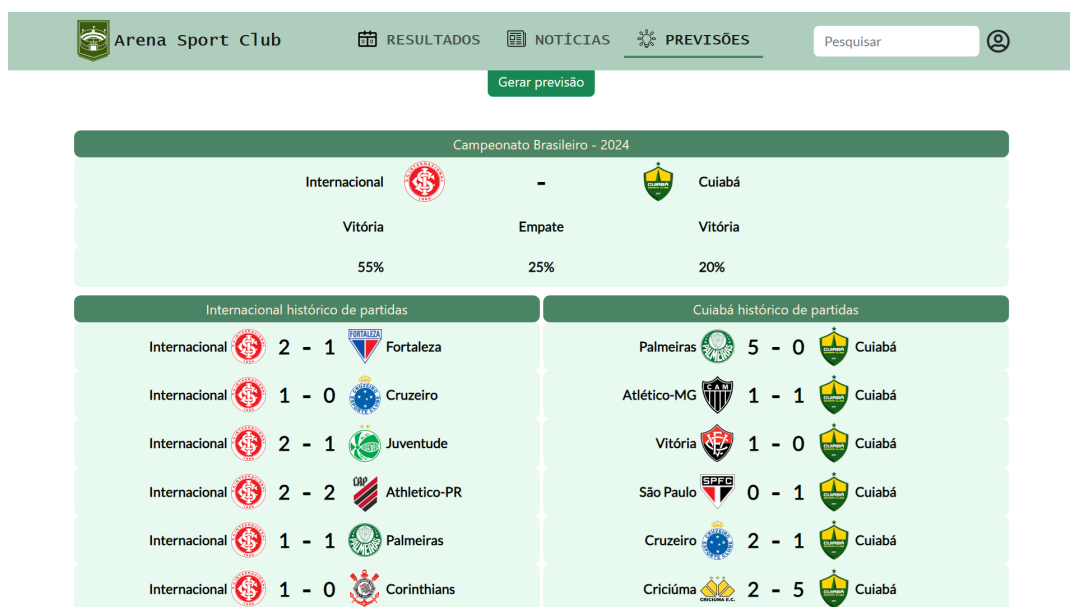


Fonte: Autor (2024)

a satisfação desses resultados, é necessário realizar testes, análises e comparações com os resultados reais, em uma amostra significativa.

Na seção seguinte, serão apresentados os testes conduzidos com o objetivo de identificar as combinações mais eficazes de variáveis e parâmetros. Além disso,

Figura 6 – Tela da funcionalidade de previsão de resultados de partidas no Arena Sport Club, após gerar a previsão de uma partida



Fonte: Autor (2024)

foram discutidos e analisados os resultados obtidos das combinações de testes que exploraram diferentes variações das opções de variáveis disponíveis na ferramenta, abordando pontos de destaque em cada combinação.

4 Resultados

Na seção anterior, foi detalhado o desenvolvimento e a implementação da ferramenta de previsão de resultados de partidas de futebol no projeto Arena Sport Club. A partir de uma cuidadosa estruturação dos *prompts* e da utilização de diferentes abordagens, incluindo cenários condicionais e exemplos, o sistema foi projetado para gerar previsões com base nas probabilidades de vitória e empate de uma partida. Nesta seção de resultados, é apresentada uma análise superficial dos testes realizados para cada categoria da ferramenta, comparando as previsões geradas pelo LLM com os resultados reais das partidas. Através dessa análise, busca-se avaliar a precisão e a eficácia das previsões, proporcionando uma visão crítica sobre a capacidade do modelo em replicar os resultados de eventos esportivos com base nos parâmetros definidos.

4.1 Estruturação dos testes

Para realizar uma análise satisfatória dos resultados dos testes, foi essencial delimitar um intervalo de partidas com correlação significativa, possibilitando a comparação com resultados reais. A escolha desse intervalo pode incluir partidas pertencentes ao mesmo campeonato e/ou temporada, garantindo, assim, que os jogos estejam relacionados ou tenham ocorrido em períodos temporalmente próximos. Adicionalmente, é necessário que esse intervalo inclua um número substancial de partidas, pois a ampliação da amostra analisada tende a aumentar a consistência e a coesão dos resultados. Esse aumento na quantidade de dados pode revelar tendências nas análises, facilitando a interpretação dos resultados obtidos.

Durante a exploração de bases de dados na internet, foram encontradas diversas fontes de dados no site Football-Data¹. Este site disponibiliza planilhas contendo informações sobre diferentes campeonatos ao redor do mundo. Cada conjunto de dados inclui as partidas realizadas no respectivo campeonato, juntamente com informações básicas e seus resultados, além de dados adicionais relevantes, como as probabilidades de vitória e empate de cada partida (também conhecidas como odds), fornecidas por diferentes casas de apostas. Ao realizar os testes, torna-se possível comparar os resultados previstos pelo Arena Sport Club com os resultados reais das partidas, bem como com as probabilidades apresentadas pelas casas de apostas.

Foi selecionada a base de dados da Premier League 2023/2024 para a realização dos testes. Este campeonato, amplamente reconhecido como o mais famoso do

¹ O site Football-Data está disponível em: football-data.co.uk

mundo (EXAME, 2024), destaca-se tanto pelo elevado valor de mercado dos clubes e da liga, quanto pelo número significativo de telespectadores. No momento do início dos testes, todas as 380 partidas do campeonato já haviam sido concluídas, permitindo que todas as partidas da temporada fossem incluídas nas análises.

Os testes foram subdivididos em diversas combinações de variáveis presentes nas categorias disponibilizadas pela ferramenta de previsão. Devido às restrições de tempo e ao custo associado a cada teste, apenas algumas dessas combinações de variáveis foram avaliadas, com o objetivo de identificar a configuração que apresentasse os melhores resultados. As combinações e variáveis escolhidas serão apresentadas na [seção 4.2](#). Para isso, foi desenvolvido um script no backend da aplicação, responsável por gerar as previsões de todas as partidas do campeonato e estruturar os resultados em um arquivo CSV. A geração de cada conjunto de testes requer apenas a inserção dos IDs das partidas a serem testadas, juntamente com as variáveis selecionadas de cada categoria.

4.2 Combinações de testes

Inicialmente, foram determinadas as combinações de variáveis que seriam utilizadas nos testes iniciais. Cada combinação selecionou diferentes variáveis presentes em cada categoria, com o objetivo de encontrar o melhor resultado dentre as combinações testadas. Como a quantidade de combinações possíveis é muito grande, foi realizada a combinação separada de cada categoria, as variáveis das demais categorias mantiveram os valores previamente utilizados, a fim de evitar mudanças drásticas de um teste para outro e à medida que cada categoria obteve sua variável otimizada, seu valor foi utilizado nos testes subsequentes.

Entre as categorias disponíveis, duas se destacam por influenciar significativamente a estrutura e a apresentação dos dados nos *prompts*. A primeira categoria refere-se à forma como o histórico das partidas é analisado, permitindo a análise dos dados históricos considerando apenas as partidas realizadas como mandante ou como visitante, ou a análise de todas as partidas do histórico, independentemente do local. Como foi apresentado anteriormente, essa categoria foi criada após analisar alguns trabalhos relacionados (PRASETIO; HARLILI, 2016)(RODRIGUES; PINTO, 2022)(Football Data, 2003) que destacavam a separação dos dados entre as partidas mandantes ou visitantes.

A segunda categoria de destaque está relacionada à estruturação e apresentação das estatísticas das partidas nos *prompts*. Nessa categoria, é possível optar por detalhar as estatísticas de cada partida individualmente ou apresentar apenas a média das estatísticas em conjunto.

Foi decidido realizar uma combinação simples das duas categorias citadas, com o objetivo de identificar a combinação que apresentasse os melhores resultados. Assim, foram conduzidos quatro testes utilizando essas combinações, enquanto todas as outras categorias mantiveram os mesmos valores em todos os testes.

O modelo selecionado para a execução dos testes foi o *GPT-3.5-turbo*, que é o modelo padrão utilizado pelo ChatGPT, além de ser o menos custoso. A opção de temperatura foi mantida em 1, valor padrão adotado pela API. Optou-se pelo *One-shot learning* na categoria de número de exemplos, a fim de equilibrar a quantidade de exemplos nos prompts e o custo dos testes, considerando que um maior número de exemplos torna os testes mais onerosos.

No que diz respeito ao número de partidas consideradas no histórico, estudos como o de ([Football Data, 2003](#)) indicam que a forma recente das equipes é geralmente avaliada com base nas últimas 4 a 6 partidas. Para os testes, foi selecionado o valor de 6 partidas no histórico como padrão. Quanto às estatísticas das partidas, a análise dos trabalhos revisados mostrou que as 7 variáveis presentes na funcionalidade de previsão são consistentemente utilizadas, elas são: posse de bola (%), total de passes, passes corretos (%), total de chutes, chutes no gol, escanteios e faltas cometidas. Dado que todas as variáveis são citadas nos trabalhos relacionados, decidiu-se incluir todas elas em todos os testes.

Partindo para a última categoria, as estatísticas do campeonato apresentam um problema específico, dado que os testes foram iniciados após o término do campeonato. As estatísticas disponibilizadas pelo FBref referem-se à última rodada da competição. Isso implica que, ao testar partidas da 1ª rodada utilizando dados do final do campeonato, as previsões seriam irrealistas, uma vez que essas informações não estariam disponíveis no momento em que as partidas ocorreram. Em vista disso, optou-se por não utilizar estatísticas de campeonato nos testes principais. No entanto, foi realizado um teste específico envolvendo a categoria de estatísticas de campeonato, cujos resultados foram apresentados na [subseção 4.2.5](#).

Para avaliar a eficácia de cada teste, os resultados foram comparados com os resultados reais das partidas e com as odds (probabilidades) médias das casas de apostas. A maior probabilidade foi considerada como a previsão escolhida pelo modelo; ou seja, se o modelo indicar que a probabilidade de vitória do time mandante é a maior, o time mandante será considerado o vitorioso pela previsão. Serão apresentados tanto o número de acertos quanto a porcentagem de acertos dos testes, comparando-os com os resultados reais e com as médias das previsões das casas de apostas.

O esperado é que os resultados dos testes fiquem próximos aos resultados das casas de apostas, o que representaria um grande sucesso ao gerar as previsões utilizando o ChatGPT, já que as casas de apostas possuem diversas técnicas e abor-

dagens complexas, utilizando uma quantidade imensa de dados. Enquanto ao utilizar o ChatGPT, apenas é necessário manipular os dados enviados junto aos *prompts* formulados e ele cuida de toda a parte de processamento e geração da previsão.

Em cada bateria de testes, será detalhado o que foi testado, seguido da apresentação e uma breve discussão dos resultados. Ao final dos testes, os resultados gerais serão apresentados, acompanhados de uma análise mais aprofundada dos dados obtidos.

4.2.1 Combinação simples entre duas categorias

Na combinação das duas categorias de destaque, foram realizados quatro testes, cada um prevendo os resultados das 380 partidas da temporada 2023/2024 da Premier League. O [Quadro 2](#) apresenta a identificação de cada teste, juntamente com as variáveis selecionadas em cada um deles. A [Tabela 1](#) exhibe os resultados das previsões, comparando-os com os resultados reais das partidas.

Quadro 2: Identificação e detalhes dos 4 testes iniciais realizados

Identificação do teste	Variáveis selecionadas
Combinação 1	Todas partidas do histórico Apresentação das estatísticas por partida
Combinação 2	Todas partidas do histórico Apresentação das estatísticas por média
Combinação 3	Apenas partidas mandante ou visitante do histórico Apresentação das estatísticas por média
Combinação 4	Apenas partidas mandante ou visitante do histórico Apresentação das estatísticas por partida

Fonte: Autor (2024)

Tabela 1 – Comparação dos resultados dos 4 testes iniciais aos resultados reais

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 1	213	56,05%
Combinação 2	205	53,95%
Combinação 3	213	56,05%
Combinação 4	218	57,37%

Fonte: Autor (2024)

Os resultados indicam que a "Combinação 4" obteve o melhor desempenho, aproximando-se dos resultados das casas de apostas. Com exceção da "Combinação

2”, todas as outras combinações apresentaram uma porcentagem de acertos superior a 55%, e os resultados entre elas foram bastante semelhantes. A partir de agora, as variáveis da “Combinação 4” serão aplicadas em todos os testes subsequentes. No [Quadro 3](#) é possível visualizar as variáveis utilizadas na “Combinação 4”.

Quadro 3: Variáveis utilizadas na “Combinação 4”

Categoria	Variáveis selecionadas
Modelo	<i>GPT-3.5-turbo</i>
Temperatura	Valor 1
Número de exemplos para o modelo	<i>One-shot</i>
Número de partidas no histórico	6 partidas
Histórico de partidas a serem analisadas	Apenas partidas mandante ou visitante do histórico
Estatísticas das partidas	Todas disponíveis
Apresentação das estatísticas	Apresentação das estatísticas por partida
Estatísticas do campeonato	Nenhuma

Fonte: Autor (2024)

4.2.2 Combinações da categoria de temperatura

A seguir, serão apresentados os resultados dos testes específicos da categoria de temperatura. Nos testes iniciais, a temperatura utilizada foi definida como 1, que é o valor padrão da API e oferece uma abordagem equilibrada entre as outras opções disponíveis. As demais opções de temperatura foram então testadas: a primeira com o valor 0, que faz com que o modelo funcione de forma mais focada e determinística, e a segunda com o valor 2, que leva o modelo a adotar um comportamento mais aleatório e criativo em suas respostas.

Ao realizar os testes utilizando o valor 2 para a temperatura, poucas respostas foram consistentes com o formato solicitado ao modelo. Em várias previsões, o modelo retornou frases extensas com caracteres aleatórios, tornando inviável a utilização desta configuração nas previsões. A [Tabela 2](#) apresenta os resultados do teste utilizando o valor de temperatura 0, comparando-os com os resultados da “Combinação 4”, que utiliza o valor de temperatura 1. Similarmente a outras combinações, o teste com “Temperatura 0” aproximou-se dos resultados da “Combinação 4”, mas ainda permaneceu ligeiramente inferior. Dessa forma, a “Combinação 4” continua a ser a mais eficaz até o momento.

4.2.3 Combinações da categoria de número de exemplos para o modelo

A próxima categoria testada foi o número de exemplos fornecidos ao modelo. Na “Combinação 4”, foi utilizada a opção de *One-shot learning*, onde o modelo recebe um exemplo antes de processar o prompt de dados. Duas outras opções foram testadas: a primeira é o *Zero-shot learning*, que não oferece nenhum exemplo ao modelo, e a segunda é o *Three-shot learning*, que fornece três exemplos distintos. A [Tabela 3](#) apresenta os resultados de cada teste realizado nesta categoria.

Mais uma vez, os resultados da “Combinação 4” se destacaram, permanecendo superiores aos novos testes. Dentro da categoria específica de número de exemplos, o teste “*Three-shots learning*” apresentou um desempenho melhor em comparação ao “*Zero-shot learning*”. Além disso, em alguns testes isolados utilizando o *Zero-shot*, as respostas do modelo mostraram-se inconsistentes, com uma estruturação diferente da solicitada, o que comprometeu a qualidade das previsões.

4.2.4 Combinações da categoria de número de partidas no histórico

A seguir, foi testada a categoria referente ao número de partidas adicionadas ao histórico. Como mencionado anteriormente, foi selecionado o valor 6 nos testes iniciais, com base no trabalho de ([Football Data, 2003](#)), que sugere que a forma recente das equipes é geralmente avaliada com base nas últimas 4 a 6 partidas. O valor 6 foi

Tabela 2 – Comparação entre os resultados reais com os resultados do teste de temperatura

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 4	218	57,37%
Temperatura 0	216	56,84%

Fonte: Autor (2024)

Tabela 3 – Comparação entre os resultados reais com os resultados dos testes de número de exemplos para o modelo

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 4	218	57,37%
<i>Zero-shot learning</i>	207	54,47%
<i>Three-shots learning</i>	212	55,79%

Fonte: Autor (2024)

escolhido também por ser um valor intermediário entre 1 e 10, equilibrando o custo computacional e a riqueza de dados históricos utilizados.

Foram testados dois valores distintos para o número de partidas no histórico. No primeiro teste, foi utilizado um valor que corresponde à metade do valor previamente selecionado. Testar valores muito próximos de 1 foi considerado ineficaz, pois adicionaria poucos dados ao prompt, potencialmente resultando em previsões mais inconsistentes. O valor 3 foi identificado como o valor mínimo aceitável para ser testado. No segundo teste, foi escolhido o maior valor possível, com o objetivo de avaliar se a adição de mais dados poderia melhorar os resultados ou se adotar uma opção mais equilibrada entre custo e quantidade de dados seria mais vantajoso a longo prazo. A [Tabela 4](#) mostra que, embora os resultados dos testes tenham se mantido próximos, eles ainda ficaram abaixo do desempenho alcançado pela “Combinação 4”.

Tabela 4 – Comparação entre os resultados reais com os resultados dos testes do número de partidas adicionadas ao histórico

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 4	218	57,37%
Histórico 3 partidas	212	55,79%
Histórico 10 partidas	210	55,26%

Fonte: Autor (2024)

4.2.5 Combinações da categoria de estatísticas do campeonato

Conforme mencionado anteriormente na sessão de combinações de testes, as estatísticas do campeonato não seriam utilizadas, uma vez que esses dados se referem apenas às informações disponíveis após o término do campeonato. No entanto, para fins de análise, foi realizada uma combinação de testes utilizando algumas dessas estatísticas. As estatísticas selecionadas incluem a posição e pontuação da equipe na tabela, partidas disputadas, vitórias, empates, derrotas, gols marcados, gols sofridos, saldo de gols, posse de bola e passes completos.

Foi realizada uma combinação de testes adicionais que utilizou dados de salários semanais das equipes, associados à idade média do elenco. Esses dados poderiam ser aplicados a qualquer partida do campeonato, pois permanecem constantes do início ao fim, desconsiderando as transferências de jogadores ao longo da temporada. Ao analisar os dados de salário médio de cada equipe, disponibilizados no site FBref, observou-se que a equipe campeã, o Manchester City, possuía uma folha salarial 7 vezes maior que a do lanterna do campeonato, o Sheffield United. Isso pode indicar

uma relação direta entre desempenho no campeonato e folha salarial, sugerindo que a adição desses dados aos prompts poderia gerar probabilidades maiores para equipes com maiores custos salariais.

Contudo, essa premissa não deve ser considerada uma regra, dado que a equipe que mais gastou com salários, o Manchester United, terminou apenas na 8ª posição no campeonato. Ao analisar os resultados na [Tabela 5](#), verifica-se que ambos os testes não conseguiram superar os resultados da “Combinação 4”.

4.2.6 Combinações da categoria do modelo

Após a realização de todas as combinações, a “Combinação 4” apresentou o melhor resultado. Com base nisso, foi realizado um novo teste, desta vez utilizando um modelo diferente. O modelo anterior utilizado era o *GPT-3.5-turbo*, e no novo teste foi utilizado o modelo *GPT-4o*. Devido ao custo elevado do modelo *GPT-4-turbo*, ele não será incorporado nos testes. Ao examinar os resultados na [Tabela 6](#), observa-se que o modelo *GPT-3.5-turbo* ainda se destacou em comparação ao *GPT-4o*, apesar de ser uma versão mais antiga e com menor capacidade de processamento simultâneo. Assim como nas outras combinações de testes, é importante ressaltar que, para alcançar uma maior precisão nos resultados, seria necessário realizar múltiplos testes com as mesmas combinações.

Tabela 5 – Comparação entre os resultados reais com os resultados dos testes das estatísticas do campeonato

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 4	218	57,37%
Dados campeonato	216	56,84%
Dados salários	211	55,53%

Fonte: Autor (2024)

Tabela 6 – Comparação entre os resultados reais com o resultado dos testes de melhor resultado em diferentes modelos

Testes realizados	Comparação com os resultados reais	
	Acertos	% de acertos
Casas de apostas	224	58,95%
Combinação 4 (<i>GPT-3.5-turbo</i>)	218	57,37%
Combinação 4 com <i>GPT-4o</i>	211	55,53%

Fonte: Autor (2024)

4.3 Comparações com os resultados gerais

Após analisar detalhadamente cada combinação de testes, explorando as diferentes opções disponíveis na funcionalidade de previsão, a [Tabela 7](#) apresenta uma visão geral dos resultados de todas as combinações realizadas. Na comparação com os resultados reais, observa-se que as odds médias das casas de apostas alcançaram uma taxa de acerto de 58,95%, o que não está muito distante do resultado obtido pela "Combinação 4" que alcançou 57,37% de acertos, o que representa apenas 6 acertos de diferença entre os dois resultados. As demais combinações mantiveram-se com resultados similares.

Tabela 7 – Comparação entre todos os testes realizados com os resultados reais e com as odds médias das casas de apostas

Testes realizados	Comparação com os resultados reais		Comparação com as odds médias das casas de apostas	
	Acertos	% de acertos	Acertos	% de acertos
Casas de apostas	224	58,95%	-	-
Combinação 1	213	56,05%	302	79,47%
Combinação 2	205	53,95%	310	81,58%
Combinação 3	213	56,05%	310	81,58%
Combinação 4	218	57,37%	312	82,11%
Temperatura 0	216	56,84%	315	82,89%
<i>Zero-shot learning</i>	207	54,47%	293	77,11%
<i>Three-shots learning</i>	212	55,79%	306	80,53%
Histórico 3 partidas	212	55,79%	308	81,05%
Histórico 10 partidas	210	55,26%	307	80,79%
Dados campeonato	216	56,84%	337	88,68%
Dados salários	211	55,53%	319	83,95%
Combinação 4 <i>GPT-4o</i>	211	55,53%	333	87,63%

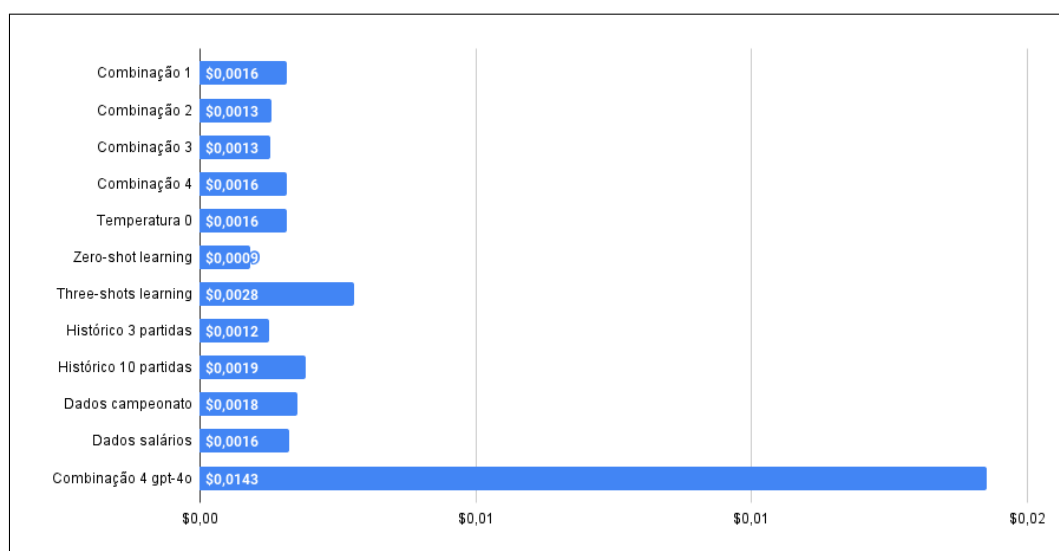
Fonte: Autor (2024)

Nesta tabela, também foi incluída uma comparação com as odds médias das casas de apostas. Em outras palavras, foram comparados o número de partidas em que a previsão do modelo e a previsão das casas de apostas indicaram o mesmo resultado final. Nesta comparação, os resultados mostram uma variação maior entre as diferentes combinações, com destaque para as combinações "Dados campeonato" e "Combinação 4 *GPT-4o*" que alcançaram, respectivamente, 88,68% e 87,63% de acertos, o que indica que quando as casas de apostas erravam a previsão real das partidas, essas combinações erravam junto, mas quando as casas de apostas acertaram a previsão, a proporção de acertos dessas combinações eram um pouco abaixo.

4.4 Custos por requisição à API do ChatGPT

Outro ponto importante na realização dos testes é o custo por requisição à API do ChatGPT. O [Gráfico 1](#) apresenta um gráfico que ilustra o custo médio por chamada da API em cada combinação de testes realizados. Observa-se que a combinação que utilizou o modelo *GPT-4o* custa quase 10 vezes mais que a média das combinações que utilizaram o modelo *GPT-3.5-turbo*. As pequenas variações entre as combinações que utilizam o *GPT-3.5-turbo* são decorrentes das alterações nas variáveis adotadas, que impactam o tamanho dos dados adicionados ao *prompt*. Por exemplo, ao utilizar o *Three-shots learning*, o custo mais que triplica em comparação ao *Zero-shot learning*. Alterar o número de partidas no histórico também provoca variações no custo médio. No final, o objetivo dos testes é identificar a combinação que ofereça o melhor equilíbrio entre custo e desempenho.

Gráfico 1: Gráfico com o custo médio por chamada da API em cada combinação de testes realizados



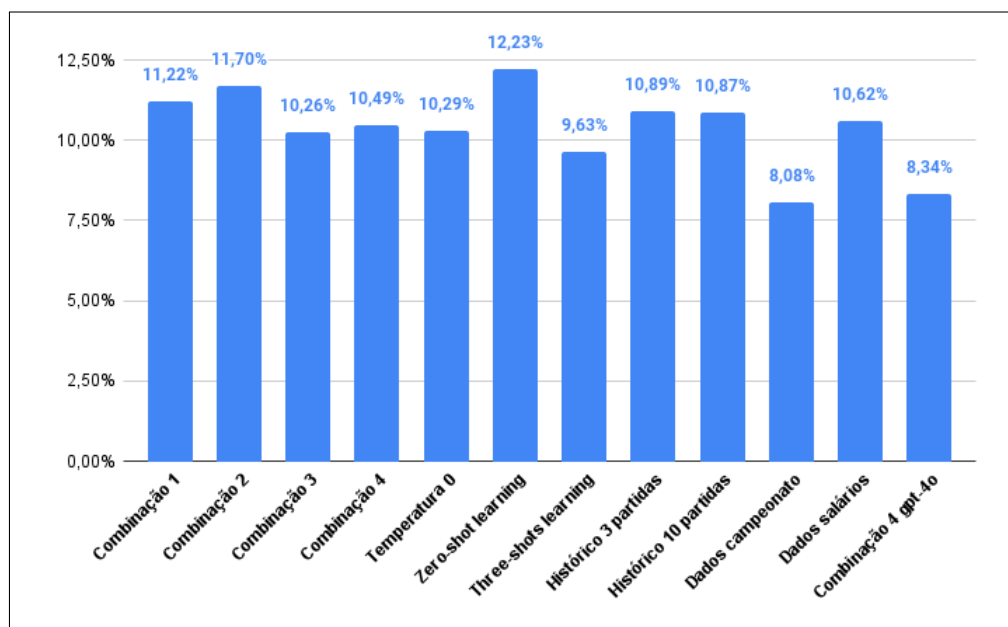
Fonte: Autor (2024)

4.5 Dispersão média dos resultados

Por fim, o [Gráfico 2](#) ilustra a dispersão média da probabilidade do vencedor da partida, comparando a média das casas de apostas com cada combinação de testes realizados. Similar aos resultados de acertos na comparação com as odds médias das casas de apostas, as combinações "Dados campeonato" e "Combinação 4 GPT-4o" já eram esperados de obterem uma dispersão mais baixa, por possuírem mais informações disponibilizadas ao *prompt* e no caso da "Combinação 4 GPT-4o" o modelo possui uma capacidade de processamento maior que o modelo *GPT-3.5-turbo*. A "Combinação 4" obteve um resultado mediano comparado aos demais testes, o que pode indicar

que apesar das alternâncias entre algumas variáveis, o modelo se mantém consistente nas previsões.

Gráfico 2: Gráfico com a dispersão média da probabilidade do vencedor da partida na média das casas de apostas com cada combinação de testes realizados



Fonte: Autor (2024)

Após a apresentação dos resultados e uma análise inicial das previsões geradas pela ferramenta presente no Arena Sport Club, a próxima seção busca realizar uma discussão mais detalhada, com o intuito de contextualizar os achados obtidos. Serão destacadas as conclusões do estudo e sugeridas direções para futuras pesquisas, visando aprimorar o sistema e expandir suas potenciais aplicações. Posteriormente, serão examinados os pontos fortes, pontos fracos e as limitações das estratégias adotadas, além de explorar como esses resultados se comparam com as previsões das casas de apostas.

5 Conclusão e discussão dos resultados

Neste trabalho, foi desenvolvida uma ferramenta capaz de gerar previsões sobre o resultado de partidas de futebol. Essa ferramenta faz parte do projeto Arena Sport Club, e sua implementação exigiu a construção de uma arquitetura robusta e a obtenção de uma vasta quantidade de dados. Esses dados foram coletados de duas fontes distintas, ambas capturadas por meio de técnicas de *Web Scraping*.

Após a coleta, os dados foram armazenados e organizados de forma estruturada. A ferramenta permite a personalização das previsões por meio da seleção de diversas opções que influenciam os dados enviados e a forma como eles são apresentados. Por fim, as solicitações são padronizadas e enviadas para a API do ChatGPT, que processa e analisa as informações, gerando uma previsão que é então apresentada ao usuário na interface da ferramenta.

Cada partida de futebol apresenta características únicas, com inúmeras variáveis que podem influenciar seu resultado. Muitas dessas variáveis não são imediatamente perceptíveis ou analisadas, o que contribui para a singularidade de cada jogo. Ao longo do tempo, as equipes e os dados referentes a elas passam por mudanças, que podem representar uma evolução ou regressão. Por essa razão, a análise cuidadosa e a interpretação dessas variáveis são fundamentais para gerar previsões precisas. Ao utilizar a ferramenta desenvolvida, é possível manipular como os dados serão utilizados e estruturados, permitindo que cada usuário adote seu próprio método analítico. Isso possibilita a geração de diferentes previsões, adaptadas à abordagem escolhida por cada usuário.

5.1 Discussão dos resultados obtidos

Os resultados apresentados na [Tabela 7](#) exibem diversas características e pontos de destaque. Em uma análise geral, torna-se difícil identificar qual combinação seria efetivamente a mais adequada, devido ao número limitado de testes realizados, o que impede uma conclusão com alto grau de confiança. No entanto, certos aspectos se sobressaem, tanto de forma positiva quanto negativa, os quais, ao serem analisados de maneira criteriosa, podem revelar características importantes e fornecer *insights* relevantes.

Inicialmente, ao comparar os resultados das casas de apostas com os resultados reais, observa-se que a taxa de acerto foi de apenas 58,95%. Esse desempenho pode ser explicado, em grande parte, pelos erros de previsão em partidas que não

foram vencidas por equipes que eram consideradas altamente favoritas à vitória. Resultados desse tipo sempre ocorreram e continuarão a ocorrer, no entanto, a frequência com que equipes altamente favoritas não conseguem alcançar a vitória varia entre cada campeonato.

Ao analisar o resultado que mais se aproximou do desempenho das casas de apostas, observado na "Combinação 4", que obteve uma taxa de acerto de 57,37%, é possível realizar comparações com trabalhos previamente analisados. Por exemplo, o estudo (RODRIGUES; PINTO, 2022) apresentou uma taxa de acerto de 65,26%, utilizando uma abordagem baseada em árvores de decisão, enquanto o trabalho (PRA-SETIO; HARLILI, 2016) atingiu uma taxa de 69,5%, ao empregar métodos estatísticos tradicionais como a regressão logística.

Comparando esses resultados com o experimento atual, observa-se que a "Combinação 4" apresentou um desempenho abaixo entre esses dois trabalhos. Apesar disso, a "Combinação 4" ainda se mostrou competitiva, aproximando-se dos resultados das casas de apostas, na comparação desta pesquisa. Isso sugere que, apesar da margem de acerto abaixo aos métodos tradicionais, as abordagens que envolvem modelos de linguagem como o utilizado nesta pesquisa oferecem uma alternativa viável, especialmente com ajustes e personalizações adicionais.

Ao comparar as combinações de testes com as previsões de casas de apostas, observa-se que a maioria dos testes apresentou resultados satisfatórios, com taxas de acerto superiores a 80%. As combinações que obtiveram os melhores resultados nessa comparação também apresentaram as menores taxas de dispersão média, o que indica uma maior consistência nas previsões. Esse comportamento sugere que, embora as previsões estejam relativamente próximas entre si, algumas combinações específicas conseguem alcançar um desempenho superior, aproximando-se mais das previsões realizadas pelas casas de apostas.

Ao analisar os testes que apresentaram uma taxa de acerto inferior a 55%, destacam-se a "Combinação 2" e o "Zero-shot learning". No caso da "Combinação 2", não é possível descartá-la de forma definitiva, uma vez que as diferenças entre as quatro primeiras combinações são mínimas. Além disso, quando comparada às ODDs médias de casas de apostas, a "Combinação 2" apresentou um desempenho mediano em relação às demais, não justificando sua exclusão com base nos resultados obtidos.

No que diz respeito ao teste "Zero-shot learning", algumas características merecem destaque. Primeiramente, ao não fornecer exemplos ao modelo sobre o que se espera como resposta, não há garantia de que o LLM manterá consistência em suas previsões. Como mostrado no Gráfico 2, esse método apresentou a maior dispersão média entre todos os testes, corroborando a falta de consistência mencionada. Em contrapartida, conforme evidenciado no Gráfico 1, o "Zero-shot learning" teve o menor

custo médio por requisição. Contudo, ao se comparar o custo com o benefício, essa combinação não é recomendada para alcançar resultados satisfatórios.

Por outro lado, a combinação "*Three-shots learning*" apresenta um custo três vezes maior do que o teste anterior, mas o aumento de desempenho nos resultados não é proporcional ao aumento do custo. Embora essa combinação possua uma das menores dispersões médias entre as demais, essa vantagem isolada não é suficiente para torná-la uma opção viável em termos de custo-benefício.

Outra combinação que apresenta características de destaque é a "Temperatura 0". Esse teste produziu resultados muito semelhantes aos da "Combinação 4", que utilizou a temperatura 1. Como a temperatura é um parâmetro numérico decimal, existem inúmeras opções de valores possíveis. Apenas por meio de uma série de testes extensivos, com a aplicação de diferentes valores de temperatura, seria possível determinar o valor ótimo para esse parâmetro.

Uma das combinações com maior potencial, mas que não pôde ser explorada em sua totalidade, foi a "Dados Campeonato". Devido às limitações na disponibilidade de informações para cada rodada do campeonato, o uso dos dados do final da competição para todas as partidas resultaria em uma análise irrealista. No entanto, deixando essa limitação de lado, os resultados obtidos foram bastante promissores. Esta combinação apresentou a menor porcentagem de dispersão média, o melhor desempenho em comparação às casas de apostas, alcançando uma taxa de acerto de 88,68%, e apenas um pequeno acréscimo no custo médio. Esses dados sugerem que, com a aplicação adequada, essa combinação pode gerar resultados altamente positivos e representar um diferencial significativo para aumentar a taxa de acerto.

Por fim, é importante destacar que o teste final utilizando o modelo *GPT-4o* não atendeu às expectativas, apresentando um desempenho inferior na comparação com os resultados reais. Considerando que este modelo possui um custo quase 10 vezes maior, esperava-se um ganho significativo de performance. Entretanto, na comparação com as casas de apostas, houve um aumento de somente 5% na taxa de acertos comparado ao modelo *GPT-3.5-turbo*. Em relação à dispersão média, o modelo obteve o segundo melhor resultado, perdendo apenas para a combinação "Dados Campeonato".

Embora os resultados tenham se aproximado mais das previsões das casas de apostas, o principal desafio permanece sendo o custo elevado do modelo. Com base nos resultados, pode-se supor que a combinação do modelo *GPT-4o* com os dados do campeonato poderia proporcionar um ganho de desempenho suficiente para justificar o alto custo. No entanto, para confirmar essa hipótese, seria necessário realizar uma extensa bateria de testes, a fim de assegurar que os resultados sejam consistentes e eficientes e essa bateria custaria mais do que qualquer outro teste feito neste trabalho.

Embora o custo seja um fator importante na escolha do modelo, conforme mencionado no estudo (BROWN et al., 2020), a precisão pode ser significativamente melhorada com a combinação adequada de dados e métodos avançados de LLMs. No entanto, o custo-benefício ainda deve ser avaliado cuidadosamente, como os resultados sugerem.

5.2 Pontos fortes, fracos e limitações dos resultados obtidos

Os resultados obtidos nos testes realizados mostram uma proximidade significativa em relação aos resultados de casas de apostas, quando comparados com os resultados reais. Vale destacar que casas de apostas utilizam diversas técnicas e cálculos complexos para gerar as probabilidades iniciais de uma partida. No caso da ferramenta desenvolvida, ao utilizar o ChatGPT para gerar as previsões, grande parte do trabalho analítico complexo e custoso é delegado ao modelo de linguagem. Esse modelo é capaz de abstrair as solicitações e processar as informações recebidas para produzir uma previsão satisfatória.

Outro aspecto relevante é que o ChatGPT já dispõe de uma vasta quantidade de dados acumulados durante seu treinamento, que são utilizados no processo de previsão. Ao formular *prompts* bem estruturados, incorporando dados recentes das partidas e combinando-os com os dados presentes no treinamento do modelo, percebe-se que o LLM demonstra grande capacidade de processar as informações e gerar previsões com alto potencial de precisão.

Sendo uma área relativamente nova em termos de popularidade social, o uso de LLMs abriu uma vasta gama de possibilidades. A facilidade de integração desses modelos em ferramentas que suportam diversos formatos, como texto, imagens e até áudios, tem impulsionado um rápido crescimento no número de aplicações. A cada dia surgem novas combinações e ideias com grande potencial de aplicação. No entanto, ainda há uma escassez de estudos que documentem dados e experiências sobre o uso desses modelos em aplicações de larga escala, evidenciando a necessidade de mais pesquisas nesse campo emergente.

No presente trabalho, foram identificadas combinações de testes que geraram resultados satisfatórios, mantendo o custo relativamente baixo. Após a coleta dos dados necessários para o funcionamento da ferramenta, a maior parte do esforço concentrou-se na estruturação dos prompts e na criação de diferentes opções para diversificar as previsões geradas. A utilização de um LLM pode ser bastante vantajosa, mas sua eficácia depende do cenário de aplicação e da viabilidade do custo envolvido. À medida que novos modelos mais poderosos e acessíveis forem desenvolvidos, a tendência é que o uso de LLMs se torne cada vez mais viável e difundido em diver-

sas aplicações, o que pode representar um ponto decisivo para sua adoção em larga escala.

Contudo, é fundamental reconhecer as limitações inerentes aos modelos de linguagem. O processo realizado por um LLM é, em grande parte, uma "caixa-preta", o que significa que não é possível determinar com precisão como o modelo processa cada solicitação internamente. Além disso, uma simples atualização do modelo pode desencadear alterações nas regras de negócio da aplicação, o que pode impactar diretamente os resultados esperados. Identificar essas mudanças não é uma tarefa trivial, o que acrescenta um grau de imprevisibilidade ao uso desses modelos.

Um outro ponto pertinente é que a utilização de sites de apostas, especialmente no contexto de previsões esportivas, levanta questões éticas relevantes, uma vez que essas plataformas estão associadas a um risco significativo de vício e comportamentos compulsivos. O fácil acesso e a gamificação das apostas podem criar um ciclo de dependência, afetando negativamente a saúde mental e financeira dos usuários. Embora alguns usuários possam interagir de maneira controlada, o ambiente das apostas online apresenta uma série de armadilhas éticas, que vão desde a falta de transparência nos algoritmos até a ausência de bloqueios efetivos para mitigar o vício em jogos de azar. Trabalhos futuros sobre este tema são essenciais para investigar as implicações éticas e sociais do uso massivo de sites de apostas, propondo regulações mais rigorosas e políticas públicas que incentivem o jogo responsável.

Em relação às limitações deste estudo, cabe destacar que, devido a restrições de tempo e recursos financeiros, testes extensivos que comprovassem a consistência dos resultados não puderam ser realizados. Além disso, novas combinações de testes e versões do ChatGPT surgiram ao longo da pesquisa, dificultando a exploração de todo o potencial dessas atualizações. Essas limitações apontam para a necessidade de pesquisas futuras que investiguem os novos modelos de linguagem e combinações de testes, com o intuito de validar e expandir os achados aqui apresentados.

Também é possível utilizar diferentes LLMs e realizar uma comparação com os resultados gerados pelo ChatGPT ou utilizar a funcionalidade de Fine-Tuning do próprio ChatGPT, que possui um grande potencial de uso. Outra opção de expansão da funcionalidade de previsão é gerar mais respostas geradas pelo modelo além das probabilidades da previsão, uma opção pode ser gerar resumos do porque essa probabilidade foi gerada. Isso pode gerar uma confiabilidade e convencimento maior ao visualizar as probabilidades geradas em conjunto.

Em síntese, este trabalho não apenas contribui para a compreensão do uso de LLMs em aplicações do mundo real, mas também abre novos caminhos para investigações futuras, reafirmando a importância de continuar explorando o tema para o progresso contínuo na área de inteligência artificial.

Referências

ANIL, R. et al. *PaLM 2 Technical Report*. 2023. Disponível em: <<https://arxiv.org/abs/2305.10403>>.

BANG, Y. et al. *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity*. 2023. Disponível em: <<https://arxiv.org/abs/2302.04023>>.

BOMMASANI, R. et al. *On the Opportunities and Risks of Foundation Models*. 2022. Disponível em: <<https://arxiv.org/abs/2108.07258>>.

BROWN, T. B. et al. *Language Models are Few-Shot Learners*. 2020. Disponível em: <<https://arxiv.org/abs/2005.14165>>.

CHEN, L.; ZAHARIA, M.; ZOU, J. *How is ChatGPT's behavior changing over time?* 2023. Disponível em: <<https://arxiv.org/abs/2307.09009>>.

DAI, Z. et al. *Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context*. 2019. Disponível em: <<https://arxiv.org/abs/1901.02860>>.

DUARTE, F. *Number of ChatGPT Users (2023)*. 2023. Acesso em: 29 jun. 2023. Disponível em: <<https://explodingtopics.com/blog/chatgpt-users>>.

EXAME. *Cada vez mais valorizada, Premier League abre temporada com maior investimento do futebol mundial*. 2024. Acesso em: 14 de setembro de 2024. Disponível em: <<https://exame.com/esporte/cada-vez-mais-valorizada-premier-league-abre-temporada-com-maior-investimento-do-futebol->>.

Football Data. *Rating Systems for Fixed Odds Football Match Predictions*. 2003. <<https://www.football-data.co.uk/ratings.pdf>>. Acesso em: 14/09/2024.

IEVOLI, R.; PALAZZO, L.; RAGOZINI, G. On the use of passing network indicators to predict football outcomes. *Knowledge-Based Systems*, v. 222, p. 106997, 2021. ISSN 0950-7051. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705121002604>>.

KENTON, J. D. M.-W. C.; TOUTANOVA, L. K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of naacL-HLT*. [S.l.: s.n.], 2019. v. 1, p. 2.

LIU, Y. et al. Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *Journal of Materiomics*, v. 9, n. 4, p. 798–816, 2023. ISSN 2352-8478. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352847823000771>>.

MAHER, M. J. Modelling association football scores. *Statistica Neerlandica*, v. 36, p. 109–118, 1982.

MASSE, M. *REST API design rulebook*. [S.l.]: "O'Reilly Media, Inc.", 2011.

- NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, v. 18, n. 5, p. 544–551, 09 2011. ISSN 1067-5027. Disponível em: <<https://doi.org/10.1136/amiajnl-2011-000464>>.
- PRASETIO, D.; HARLILI, D. Predicting football match results with logistic regression. In: *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*. [S.l.: s.n.], 2016. p. 1–5.
- RADFORD, A. Improving language understanding by generative pre-training. 2018.
- RODRIGUES, F.; PINTO Ângelo. Prediction of football match results with machine learning. *Procedia Computer Science*, v. 204, p. 463–470, 2022. ISSN 1877-0509. International Conference on Industry Sciences and Computer Science Innovation. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050922007955>>.
- SHIN, J.; GASPARYAN, R. A novel way to soccer match prediction. *Stanford University: Department of Computer Science*, 2014.
- TOUVRON, H. et al. *Llama 2: Open Foundation and Fine-Tuned Chat Models*. 2023. Disponível em: <<https://arxiv.org/abs/2307.09288>>.
- VASWANI, A. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.
- WHITE, J. et al. *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. 2023. Disponível em: <<https://arxiv.org/abs/2302.11382>>.
- WHITE, J. et al. Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. In: _____. *Generative AI for Effective Software Development*. Cham: Springer Nature Switzerland, 2024. p. 71–108. ISBN 978-3-031-55642-5. Disponível em: <https://doi.org/10.1007/978-3-031-55642-5_4>.
- WU, Y. chen; FENG, J. wen. Development and application of artificial neural network. *Wireless Personal Communications*, v. 102, n. 2, p. 1645–1656, 2018. ISSN 1572-834X. Disponível em: <<https://doi.org/10.1007/s11277-017-5224-x>>.
- XUE, L. mt5: A massively multilingual pre-trained text-to-text transformer. *arXiv preprint arXiv:2010.11934*, 2020.
- ZHAO, B. Web scraping. *Encyclopedia of big data*, Springer New York, v. 1, 2017.